

Isoform 1:

```

1  TTGCTCACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT
51 CCTCTTCCTT CTGGGGGTCC TGGGGGCCCT CACTGAGATG TGTGAAATAC
101 CAGAGATGGA CAGCCATCTG GTAGAGAAGT TGGGCCAGCA CCTCTTACCT
151 TGGATGGACC GGCTTTCCCT GGAGCACTTG AACCCAGCA TCTATGTGGG
201 CCTACGCCTC TCCAGTCTGC AGGCTGGGAC CAAGGAAGAC CTCTACCTGC
251 ACAGCCTCAA GCTTGGTTAC CAGCAGTGCC TCCTAGGGTC TGCCTTCAGC
301 GAGGATGACG GTGACTGCCA GGGCAAGCCT TCCATGGGCC AGCTGGCCCT
351 CTACCTGCCT GCTCTCAGAG CCAACTGTGA GTTTGTCAGG GGCCACAAGG
401 GGGACAGGCT GGTCTCACAG CTCAAATGGT TCCTGGAGGA TGAGAAGAGA
451 GCCATTGACA CAGCAGCCAT GGCAGGCTTG GCATTACCT GTCTGAAGCG
501 CTCAAACCTT AACCCTGGTC GGAGACAACG GATCACCATG GCCATCAGAA
551 CAGTGCGAGA GGAGATCTTG AAGGCCAGCA CCCCCGAGG CCACTTTGGG
601 AATGTCTACA GCACCCCAT TGCATTACAG TTCCTCATGA CTTCCCCCAT
651 GCGTGGGGCA GAACTGGGAA CAGCATGTCT CAAGGCGAGG GTTGCTTTGC
701 TGGCCAGTCT GCAGGATGGA GCCTTCCAGA ATGCTCTCAT GATTTCCCAG
751 CTGCTGCCCC TTCTGAACCA CAAGACCTAC ATTGATCTGA TCTTCCAGA
801 CTGTCTGGCA CCACGAGTCA TGTTGGAACC AGCTGCTGAG ACCATTCTCT
851 AGACCCAAGA GATCATCAGT GTCACGCTGC AGGTGCTTAG TCTCTTGCCG
901 CCGTACAGAC AGTCCATCTC TGTCTGGCC GGGTCCACCG TGGAAGATGT
951 CCTGAAGAAG GCCCATGAGT TAGGAGGATT CACATATGAA ACACAGGCCT
1001 CCTTGTCAGG CCCCTACTTA ACCTCCGTGA TGGGAAAGC GGCCGGAGAA
1051 AGGGAGTTCT GGCAGCTTCT CCGAGACCCC AACACCCAC TGTGCAAGG
1101 TATTGCTGAC TACAGACCCA AGGATGGAGA AACCATTGAG CTGAGGCTGG
1151 TTAGCTGGTA GCCCCTGAGC TCCCTCATCC CAGCAGCCTC GCACACTCCC
1201 TAGGCTTCTA CCCTCCCTCC TGATGTCCCT GGAACAGGAA CTCGCCTGAC
1251 CTGTCTGTGC CCTCTGTGCA ACTTTGAGCA ATGCCCCCTG GGATCACCCC
1301 AGCCACAAGC CCTTCGAGGG CCTTATACCA TGGCCACCT TGGAGCAGAG
1351 AGCCAAGCAT CTTCCCTGGG AAGTCTTTCT GGCCAAGTCT GGCCAGCCTG
1401 GCCCTGCAGG TCTCCCATGA AGGCCACCCC ATGGTCTGAT GGGCATGAAG
1451 CATCTCAGAC TCCTTGGAAC AAAACGGAGT CCGCAGGCCG CAGGTGTTGT
1501 GAAGACCACT CGTTCTGTGG TTGGGGTCCT GCAAGAAGGC CTCCTCAGCC
1551 CGGGGGCTAT GGCCCTGACC CCAGCTCTCC ACTCTGCTGT TAGAGTGGCA
1601 GCTCCGAGCT GGTTGTGGCA CAGTAGCTGG GGAGACCTCA GCAGGGCTGC
1651 TCAGTGCCCTG CCTCTGACAA AATTAAAGCA TTGATGGCCT GTGAAAAAAA
1701 AAAAAAAAAA AAAAAAAAAA AA

```

(SEQ. ID NO:1)

#### FEATURES:

5'UTR: 1 - 30  
 Start Codon: 31  
 Stop Codon: 1159  
 3'UTR: 1162

#### Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 108000024653390 /altid=gi 12742775 /def=ref XP_009922.2  tr...	752	0.0
CRA 108000024636236 /altid=gi 298316 /def=gb AAB25526.1  transc...	732	0.0
CRA 18000004926133 /altid=gi 339205 /def=gb AAA61057.1  (L02648...	732	0.0
CRA 108000024042036 /altid=gi 12654675 /def=gb AAH01176.1 AAH01...	731	0.0
CRA 18000004926130 /altid=gi 4507409 /def=ref NP_000346.1  tran...	727	0.0
CRA 18000004926132 /altid=gi 339203 /def=gb AAA61056.1  (L02647...	725	0.0
CRA 18000005170902 /altid=gi 7657639 /def=ref NP_056564.1  tran...	515	e-145
CRA 18000005218941 /altid=gi 4572454 /def=gb AAD23829.1 AF12128...	501	e-140
CRA 164000136745249 /altid=gi 11968124 /def=ref NP_071979.1  tr...	481	e-134
CRA 18000004926134 /altid=gi 4507407 /def=ref NP_001053.1  tran...	108	2e-22

FIGURE 1A

EST:

gi 10725490 /dataset=dbest /taxon=96...	858	0.0
gi 10947399 /dataset=dbest /taxon=96...	846	0.0
gi 9121897 /dataset=dbest /taxon=9606...	846	0.0
gi 13280819 /dataset=dbest /taxon=96...	846	0.0
gi 13287907 /dataset=dbest /taxon=96...	833	0.0
gi 13286505 /dataset=dbest /taxon=96...	831	0.0
gi 8150776 /dataset=dbest /taxon=960...	815	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	726	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0

**EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

gi 10725490	adult adrenal gland
gi 10947399	mammary gland
gi 9121897	retinoblastoma
gi 13280819	adenocarcinoma cell line
gi 13287907	retinoblastoma
gi 13286505	embryonal carcinoma, cell line
gi 8150776	adult uterus
gi 5936410	adult uterus
gi 6888875	adult head_neck
gi 6888872	adult head_neck

Tissue Expression:

Human leukocyte

FIGURE 1B

Isoform 2:

```

1  GGAGGATTAA TCAGTGACAG GAAGCTGCGT CTCTCGGAGC GGTGACCAGC
51 TGTGGTCAGG AGAGCCTCAG CAGGGCCAGC CCCAGGAGTC TTTCCCGATT
101 CTTGCTCACT GCTCACCCAC CTGCTGCTGC CATGAGGCAC CTTGGGGCCT
151 TCCTCTTCCT TCTGGGGGTC CTGGGGGCCC TCACTGAGAT GTGTGAAATA
201 CCAGAGATGG ACAGCCATCT GGTAGAGAAG TTGGGCCAGC ACCTCTTACC
251 TTGGATGGAC CGGCTTTCCC TGGAGCACTT GAACCCCAGC ATCTATGTGG
301 GCCTACGCCT CTCCAGTCTG CAGGCTGGGA CCAAGGAAGA CCTCTACCTG
351 CACAGCCTCA TGCTTGGTTA CCAGCAGTGC CTCCTAGGGT CTGCCTTCAG
401 CGAGGATGAC GGTGACTGCC AGGGCAAGCC TTCCATGGGC CAGCTGGCCC
451 TCTACCTGCT CGCTCTCAGA GCCAACTGGC ATGATCACAA GGGCCACCCC
501 CACACTAGCT ACTACCAGTA TGGCCTGGGC ATTCTGGCCC TGTGTCTCCA
551 CCAGAAGCGG GTCCATGACA GCGTGGTGGA CAAACTTCTG TATGCTGTGG
601 AACCTTTCCA CCAGGGCCAC CATTCTGTGG ACACAGCAGC CATGGCAGGC
651 TTGGCATTC A CTTGTCTGAA GCGCTCAAAC TTCAACCCTG GTCGGAGACA
701 ACGGATCACC ATGGCCATCA GAACAGTGC AGAGGAGATC TTGAAGGCCC
751 AGACCCCGA GGGCCACTTT GGAATGTCT ACAGACCCC ATTGGCATT A
801 CAGTTCCTCA TGACTTCCCC CATGCGTGGG GCAGAACTGG GAACAGCATG
851 TCTCAAGGCG AGGGTTGCTT TGCTGGCCAG TCTGCAGGAT GGAGCCTTCC
901 AGAATGCTCT CATGATTTCC CAGCTGCTGC CCGTTCTGAA CCACAAGACC
951 TACATTGATC TGATCTTCCC AGACTGTCTG GCACCACGAG TCATGTTGGA
1001 ACCAGCTGCT GAGACCATT C TCAGACCCA AGAGATCATC AGTGTACGC
1051 TGCAGGTGCT TAGTCTCTTG CCGCCGTACA GACAGTCCAT CTCTGTTCTG
1101 GCCGGGTCCA CCGTGGAAGA TGTCTGAAG AAGGCCCATG AGTTAGGAGG
1151 ATTCACATAT GAAACACAGG CCTCCTTGTC AGGCCCTAC TTAACCTCCG
1201 TGATGGGGAA AGCGGCCGGA GAAAGGGAGT TCTGGCAGCT TCTCCGAGAC
1251 CCCAACACCC CACTGTTGCA AGGTATTGCT GACTACAGAC CCAAGGATGG
1301 AGAAACCATT GAGCTGAGGC TGTTAGCTG GTAGCCCCTG AGCTCCCTCA
1351 TCCCAGCAGC CTCGCACACT CCCTAGGCTT CTACCCTCCC TCCTGATGTC
1401 CCTGGAACAG GAACTCGCCT GACCCTGCTG CCACCTCCTG TGCACCTTGA
1451 GCAATGCCCC CTGGGATCAC CCCAGCCACA AGCCCTTCGA GGGCCCTATA
1501 CCATGGCCCA CTTGGAGCA GAGAGCCAAG CATCTTCCCT GGAAGTCTT
1551 TCTGGCCAAG TCTGGCCAGC CTGGCCCTGC AGGTCTCCCA TGAAGGCCAC
1601 CCCATGGTCT GATGGGCATG AAGCATCTCA GACTCCTTGG CAAAAACGG
1651 AGTCCGCAGG CCGCAGGTGT TGTGAAGACC ACTCGTTCTG TGGTTGGGGT
1701 CCTGCAAGAA GGCCTCCTCA GCGCGGGGTC TATGGCCCTG ACCCCAGCTC
1751 TCCACTCTGC TGTTAGAGTG GCAGCTCCGA GCTGGTTGTG GCACAGTAGC
1801 TGGGGAGACC TCAGCAGGGC TGCTCAGTGC CTGCCTCTGA CAAAATTAAA
1851 GCATTGATGG CCTGTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA

```

(SEQ ID NO:2)

#### FEATURES:

5'UTR: 1 - 131  
 Start Codon: 132  
 Stop Codon: 1332  
 3'UTR: 1335

#### Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 108000024636236 /altid=gi 298316 /def=gb AAB25526.1  transc...	793	0.0
CRA 108000024653390 /altid=gi 12742775 /def=ref XP_009922.2  tr...	793	0.0
CRA 18000004926133 /altid=gi 339205 /def=gb AAA61057.1  {L02648...	792	0.0
CRA 108000024042036 /altid=gi 12654675 /def=gb AAH01176.1 AAH01...	792	0.0
CRA 18000004926130 /altid=gi 4507409 /def=ref NP_000346.1  tran...	788	0.0
CRA 18000004926132 /altid=gi 339203 /def=gb AAA61056.1  {L02647...	786	0.0
CRA 18000005170902 /altid=gi 7657639 /def=ref NP_056564.1  tran...	561	e-159

FIGURE 1C

CRA 164000136745249 /altid=gi 11968124 /def=ref NP_071979.1  tr...	554	e-156
CRA 18000005218941 /altid=gi 4572454 /def=gb AAD23829.1 AF12128...	545	e-154
CRA 18000004926134 /altid=gi 4507407 /def=ref NP_001053.1  tran...	128	1e-28

**EST:**

gi 10725490 /dataset=dbest /taxon=96...	858	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	835	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0
gi 12258937 /dataset=dbest /taxon=960...	686	0.0
gi 10947399 /dataset=dbest /taxon=96...	680	0.0
gi 13287907 /dataset=dbest /taxon=96...	680	0.0
gi 9121897 /dataset=dbest /taxon=9606...	680	0.0
gi 13280819 /dataset=dbest /taxon=96...	680	0.0
gi 8150776 /dataset=dbest /taxon=960...	656	0.0

**EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

gi 10725490	adult adrenal gland
gi 5936410	adult uterus
gi 6888875	adult head_neck
gi 6888872	adult head_neck
gi 12258937	adult lung_tumor
gi 10947399	mammary gland
gi 13287907	retinoblastoma
gi 9121897	retinoblastoma
gi 13280819	adenocarcinoma cell line
gi 8150776	

Tissue Expression:

Human hippocampus

FIGURE 1D

Isoform 1:

```

1 MRHLGAFLEFL LGVLGALTEM CEIPEMDSHL VEKLGQHLLP WMDRLSLEHL
51 NPSIYVGLRL SSLQAGTKED LYLHSLKLG Y QQCLLGS AFS EDDGDCQK
101 SMGQLALYLL ALRANCEFVR GHKGDR LVSQ LKWFLEDEKR AIDTAAMAGL
151 AFTCLKRSNF NPGRRQRITM AIRTVREEIL KAQTPEGHFG NVYSTPLALQ
201 FLMTSPMRGA ELGTACLKAR VALLASLQDG AFQNALMISQ LLPVLNHKTY
251 IDLIFPDCLA PRVMLEPAAE TIPQTQEIIS VTLQVLSLLP PYRQSI SVLA
301 GSTVEDVLKK AHELGGFTYE TQASLSGPYL TSVMGKAAGE REFWQLLRDP
351 NTPLLQGIAD YRPKDGETIE LRLVSW

```

(SEQ ID NO:3)

#### FEATURES:

##### Functional domains and key regions:

PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site

Number of matches: 2

1	75-77	SLK
2	174-176	TVR

PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site

Number of matches: 6

1	67-70	TKED
2	90-93	SEDD
3	174-177	TVRE
4	226-229	SLQD
5	249-252	TYID
6	302-305	STVE

PDOC00008 PS00008 MYRISTYL  
N-myristoylation site

Number of matches: 7

1	12-17	GVLGAL
2	57-62	GLRLSS
3	86-91	GSAFSE
4	149-154	GLAFTC
5	190-195	GNVYST
6	209-214	GAELGT
7	230-235	GAFQNA

PDOC00009 PS00009 AMIDATION  
Amidation site

162-165	PGRR
---------	------

#### SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

#### **BLAST Alignment to Top Hit:**

```

>CRA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|
transcobalamin II, TC II [human, endothelial cells,
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa
/length=427
Length = 427

```

Score = 732 bits (1870), Expect = 0.0  
Identities = 376/427 (88%), Positives = 376/427 (88%), Gaps = 51/427 (11%)  
Frame = +1

## FIGURE 2A

Query: 31 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL 210  
 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL  
 Sbjct: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL 60

Query: 211 SSLQAGTKEDLYLHSLKLGYYQQCLLGSASFSEDDGDCQGKPSMGQLALYLLALRANCEFVR 390  
 SSLQAGTKEDLYLHSLKLGYYQQCLLGSASFSEDDGDCQGKPSMGQLALYLLALRANCEFVR  
 Sbjct: 61 SSLQAGTKEDLYLHSLKLGYYQQCLLGSASFSEDDGDCQGKPSMGQLALYLLALRANCEFVR 120

Query: 391 GHKGDRLVSQLKWFLDEKRAI----- 456  
 GHKGDRLVSQLKWFLDEKRAI  
 Sbjct: 121 GHKGDRLVSQLKWFLDEKRAIGHDHKGHPHTSYYYQYGLGILALCLHQKRVHDSVVDKLL 180

Query: 457 -----DTAAMAGLAFTCLKRSNFNPGRRQRITMAIRTVREEILKAQTPEGHF 597  
 DTAAMAGLAFTCLKRSNFNPGRRQRITMAIRTVREEILKAQTPEGHF  
 Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFNPGRRQRITMAIRTVREEILKAQTPEGHF 240

Query: 598 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFAQNALMISQLLPVLNHKT 777  
 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFAQNALMISQLLPVLNHKT  
 Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFAQNALMISQLLPVLNHKT 300

Query: 778 YIDLIFPDCLAPRVMLEPAAETIPQTQEIISVTLQVLSLLPPYRQSSISVLAGSTVEDVLK 957  
 YIDLIFPDCLAPRVMLEPAAETIPQTQEIISVTLQVLSLLPPYRQSSISVLAGSTVEDVLK  
 Sbjct: 301 YIDLIFPDCLAPRVMLEPAAETIPQTQEIISVTLQVLSLLPPYRQSSISVLAGSTVEDVLK 360

Query: 958 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 1137  
 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI  
 Sbjct: 361 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 420

Query: 1138 ELRLVSW 1158  
 ELRLVSW  
 Sbjct: 421 ELRLVSW 427  
 (SEQ ID NO:6)

#### HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	829.9	8.6e-246	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

#### Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [. . .]	1	11 [. . .]	3.2	2.9
PF01122	1/2	1	142 [. . .]	1	143 [. . .]	296.0	4.6e-85
PF01122	2/2	143	376 .]	197	450 .]	531.8	4.8e-156

FIGURE 2B

Isoform 2:

1 MRHLGAFLEFL LGVLGALTEM CEIPEMDSHL VEKLGQHLLP WMDRLSLEHL  
51 NPSIYVGLRL SSLQAGTKED LYLHSLMLGY QQCLLGSASF EDDGDCQGKP  
101 SMGQLALYLL ALRANWHDHK GHPHTSYQY GLGILALCLH QKRVHDSVVD  
151 KLLYAVEPFH QGHHSVDTA MAGLAFTCLK RSNFNPGRRQ RITMAIRTVR  
201 EEILKAQTPE GHFGNVYSTP LALQFLMTSP MRGAELGTAC LKARVALLAS  
251 LQDGAFQNAL MISQLLPVLN HKTYIDLIFP DCLAPRVMLE PAAETIPQTQ  
301 EIISVTLQVL SLLPPYRQSI SVLAGSTVED VLKKAHELGG FTYETQASLS  
351 GPYLTSVMGK AAGEREFWQL LRDNPNTPLLQ GIADYRPKDG ETIELRLVSW  
(SEQ ID NO:4)

**FEATURES:**

**Functional domains and key regions:**

PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site  
198-200 TVR

PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site  
Number of matches: 7

1	67-70	TKED
2	90-93	SEDD
3	147-150	SVVD
4	198-201	TVRE
5	250-253	SLQD
6	273-276	TYID
7	326-329	STVE

PDOC00008 PS00008 MYRISTYL  
N-myristoylation site  
Number of matches: 7

1	12-17	GVLGAL
2	57-62	GLRLSS
3	86-91	GSAFSE
4	173-178	GLAFTC
5	214-219	GNVYST
6	233-238	GAELGT
7	254-259	GAFQNA

PDOC00009 PS00009 AMIDATION  
Amidation site  
186-189 PGRR

PDOC00428 PS00468 COBALAMIN\_BINDING  
Eukaryotic cobalamin-binding proteins signature  
165-178 SVDTAAMAGLAFTC

**SignalP results:**

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

**BLAST Alignment to Top Hit:**

>CRA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|  
transcobalamin II, TC II [human, endothelial cells,  
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa  
/length=427  
Length = 427

FIGURE 2C

Score = 793 bits (2026), Expect = 0.0  
Identities = 399/427 (93%), Positives = 399/427 (93%), Gaps = 27/427 (6%)

Query: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60  
MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL  
Sbjct: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60

Query: 61 SSLQAGTKEDLYLHSLMLGYQQCLLGSFAFSEDDGDCQGKPSMGQLALYLLALRAN----- 115  
SSLQAGTKEDLYLHSL LGYQQCLLGSFAFSEDDGDCQGKPSMGQLALYLLALRAN  
Sbjct: 61 SSLQAGTKEDLYLHSLKLGYYQQCLLGSFAFSEDDGDCQGKPSMGQLALYLLALRANCEFVR 120

Query: 116 -----W-----HDHKGHPHTSYQYGLGILALCLHQKRVHDSVVDKLL 153  
W HDHKGHPHTSYQYGLGILALCLHQKRVHDSVVDKLL  
Sbjct: 121 GHKGDRLVSQLKWFLEDEKRAIGHDHKGHPHTSYQYGLGILALCLHQKRVHDSVVDKLL 180

Query: 154 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNPNPGRQRITMAIRTVREEILKAQTPEGHF 213  
YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNPNPGRQRITMAIRTVREEILKAQTPEGHF  
Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNPNPGRQRITMAIRTVREEILKAQTPEGHF 240

Query: 214 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAQNALMISQLLPVLNHKT 273  
GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAQNALMISQLLPVLNHKT  
Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAQNALMISQLLPVLNHKT 300

Query: 274 YIDLIFPDCLAPRVMLEPAEAETIPQTQEIIISVTLQVLSLLPPYRQSSISVLAGSTVEDVLK 333  
YIDLIFPDCLAPRVMLEPAEAETIPQTQEIIISVTLQVLSLLPPYRQSSISVLAGSTVEDVLK  
Sbjct: 301 YIDLIFPDCLAPRVMLEPAEAETIPQTQEIIISVTLQVLSLLPPYRQSSISVLAGSTVEDVLK 360

Query: 334 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 393  
KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI  
Sbjct: 361 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 420

Query: 394 ELRLVSW 400  
ELRLVSW  
Sbjct: 421 ELRLVSW 427  
(SEQ ID NO:7)

#### HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	906.3	8.6e-269	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

#### Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.]	1	11 [.]	3.2	2.9
PF01122	1/2	1	115 [.]	1	115 [.]	241.3	1.4e-68
PF01122	2/2	117	400 .]	145	450 .]	660.5	8.7e-195

FIGURE 2D



```

1 ATATGTATGG GAAATATGCT GTCTTCCTAT TCCTACTCCC CCACCCCTCTA
51 GCACTGAGTC CAGGTAGGTA GGCAGGGGGG TGCTCCTCTC CTTTACTTCG
101 ACACCCCTAAC TACC'TTGGGG ATCAGAAAGT ACTCTCTGGA AGGATGCTGC
151 TGCTTCTCAC CAGAGGCTGA CGATAACGAA GGCTATCCTC CATGGCCACC
201 TCCTCCAGGC TGCC'TTCTCTG GAAATAGGAA TCATAATAGT TGT'TACTGGA
251 AACAGGCAGA GGG'TTGGGGG AGCCAAGGCA GTCCCACCCA GGACCAAGGT
301 GGCTCCATTG CACACACTTC ACCATGACTC CCCTGAAGGT CCAAACGTGC
351 GGT'TCTGCGG AAG'TTGGGCT CCCC'ACTGGC CTCCCTCCTT CCTCAGAACC
401 TCCAGGGGTG CTCCTCCTAG TGGCCACATC CAGCCTTTCT GACTGGACAA
451 CCTATCATTT AAAAT'TTTC AGTAGTCCG TAAACAGACA CACGT'TGCTG
501 TATTTATTTA TGTCAAGGC TTGG'TTGTG ATAAGTCAGG CTCAAAAAGA
551 TTGTCTTAAA AGAGTGAACC TTGGCAATTT ACCATAAAAT AATTGCAATG
601 CAGATTGTGC ATGGAATGA TTGGAGATAT TTTAAGGTCA TAGTGTCTTC
651 ACAAATTGAG CTGAAAGGGA ACTGTTAGGA TGATCTTGCC TAACCCCTCTC
701 ATCTCACACA GGAAGAACTA TTTTAAATC GAGAGGTAA GTGACCTGGC
751 CAAAGTCACA CAGCCACCAC TAGTTAACTC GTATACATTG ATTCTCCTGT
801 GGGGCTGGG AGATGAGGAA TCTTTGTTC TCTTCCCTGT TTGCAGAGAT
851 TTTTTTTGAG GTTACTTTCC GAGT'TCTGGC AAGTACCCCT GCTTCTGGTA
901 GCTTTCATG TCGATTCAAT CTCAT'TCTTT TTATTTTATT TTATTTTGA
951 GACAGGGTCT CACTTTGTCA CCCAAGCTGG AGTGCAGTGG TGTAACTTTG
1001 GCTCACTGTA GCCTCCACCT CTTGGGTTC ACGATCCTC CTGCCTCAGC
1051 CCCCCAAGTA GCTGGGATTA CAGACGTCTG CCACCACGCC AGGCTAATTT
1101 ATGGT'TTTT GTATGTGTTT TTTGTGTTT TGTAGAGACA GTGTTTCCCC
1151 ATGTTGCCCA GGCTGGTCTC CAACTCCTGA GCTCAAGTGA TCTGCCCGCC
1201 TCAGCCTTTC AAAGTGCTAG GATTACAGGT GTGAGCCACC GTGCCCGGAC
1251 TTAATCCCAT TCTTTAACTT GTTTTGT'TTT GTCTCTCCA GGAGGCTCCC
1301 AGCCCTTTTC GATTGGTTGA GAAAAGTGGC CTGGCTGGTC TGGGGCCAGC
1351 AGCACCCACC CTCCCTCAA TTGCCCAACT CCCCCCCCCA CCGAACTGCC
1401 CAACTCCCC TCCCCAACTG CCCAACTCCC CCACCCCCAC AATCCCTCC
1451 CGCCACAAC TAGGGAGGCG GTGCTGAAAA ACAGCTGACT CCAGCAATGC
1501 TGCTCACGTG ACCACTGCAG CTGCAGCTCC CGTTCACATC CTTGTCTTGG
1551 GTAGGTGGG CACTACCAGG GGCTCCTTTG GTAAGGAGTA CCGGGTAGGC
1601 ACCCGGTCTC GCCAATCCAC CACTGGAACA GCTGGGGGGA CAGCAGACAG
1651 GCACGGTCGG ACAGACTTGA CAGATCAGGC ATCAGGCCCT CTGCCTGGT
1701 CCCGGGCTCT TTAAGCAGGA ACGTGAATGG CCTCAAGATG TCTCATATGG
1751 TCCCACTAG CTTCTCCTC CTTTGTTCCT CTACCTCAG GAGGGCTGCT
1801 CTGCCCTTCC TTCTCTGTG CTTTGGCCTT ATGTTCCCCG CCACCACAGG
1851 CTTTCCCCCG CCCCACCCCT CTGCAGACTT AGCCGTGCAT TGCAGGCATG
1901 GAGGATTAAT CAGTGACAGG AAGCTGCGTC TCTCGGAGCG GTGACCAGCT
1951 GTGGTCAGGA GAGCCTCAGC AGGGCCAGCC CCAGGAGTCT TTCCCGATTCT
2001 TTGCTCACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT
2051 CCTCTTCTCT CTGGGGGTCC TGGGGGCCCT CACTGAGATG TGTGGTGAGT
2101 AACTCGCCTC TATCCTGTGC CTCTTTCCTC CTGGGTCTT AGTGGGGTGG
2151 CTAGGGCATA GATGAGGGA ACTTACCTGC CCTTCTAAGC TCCCATAGCA
2201 GTTTGGGCTT AGCTGGACCT CAGCATTTAA CACATCCTAT TGTGATTGAT
2251 TATATGTTTG ACTCCTCACC AGACAAGATC TCCGTTAATT CAGTCATTCTG
2301 TTCACACATT CATTACGCGC ATACTGAGCC TTTTCTGTGT CAGGCCCAGT
2351 GTTAGCCTTT GGGGAACGTG CAAAGCATGA GACAAGTCTA ATCCCTGCCA
2401 TCCTAGAGCT TATGTTCTAG GGAAGGGGGA CAGACAAAAG AAATGGTTAG
2451 GTGCTCCAC CTGAAATCTC AGCATTTTGG AAGGCTGAGG CGGGAGGGGA
2501 GGATCGCTTG AGCTCAACAG TTCAAGGTCA GCCTGGGCAA CATAGGGAGA
2551 CCCCATCTCT ACAAAAAATA AAAAAAATTA AAAAAATAGT GGGCATGGGG
2601 AAGACTTTCT GAAGACCAAG AGGACACATG GGAGCTGAAA CTCGAAGGAA
2651 GAAAAGGAGC TGGCAGGAAA GGAGTGGGGG ACACACATTC TAGGCAGCAG
2701 GAAGTGAGCC TTCGGAGGTC CTGCC'TGCTC CAGCTCTGTG CCCC'AAGGGG
2751 TCTCTTGGAG CACAGTCTCC TGGGACCTGT CTATGAGTCT GAGCTTAGAG
2801 GCTCAGGGCT GCTCCTTCAG ACAGGAGGCA GAAGGCAGAC TTTGGGAAC
2851 TTGGGCGGCC CACGCGCCTT TTCTCCTCCT CTGCACCTAG GATTACGTTG
2901 AGCAATACAC TTTCACCCCT ATGGTCTCTT GAGACCCTGG GGAACCCTG
2951 AGAGGTGGGT CCAGTCAATG CCAGGTGTCA AGTGAAGAAG TCGAGGGTTG
3001 GAGGGGCTGA GTGACCCACT CAGGGTGCTC CACCTTTTCC AGAGCTTTGC
3051 TGAAC'TAGT TTTTAGAACT TGAAGCCTCG TTTGTTTTCG TTTTGT'TTTT
3101 TGT'TGAGAGA GGT'TCTCCCT CTGTTGCCCA GGCTGGAGTG CAGTGGCAGC
3151 ATCTTGGCTC ACTGCAGCCT CTGCC'TGTG GGTTCAGTG ATTCCCCAC
3201 CTCAGCCTCC CAAGTAGCTG GAGACTGCAT GTGCATACTA CCATGCTTGG
3251 CTAATTTTGT TATTTT'TTG TAGAGACAGG GTTTCGCCAT GTTGGCCAGG
3301 CTGGTCTCGA ACTCCTGGGC TCAAGTGAAG CTCTTGCCCT GGCCTCCCAA
3351 ATTGCTGAGA TTACAGGCGT GAGCCACCGT GCCCGGCCAG AACTCCAAGC
3401 CTCTCATCTG TGTTCATAA ATGCAATCAG ACACCTCAGG TCTGGGCCCA

```

FIGURE 3A

3451 GGAACCCAG CTCTTGGTTC ATGTCCGGAC AGTCCCCAGG GGAGTTCTGG  
3501 GTTCAACCAG CAAGAGCTCT TCCTCCTGGC TGATCTGGTC CTCAGCCTTG  
3551 GACAGTTAGT CCATTAACTT GACCCACAG GAGCCCCAAT CCCTTGGGGT  
3601 CTGGGGAATC TTGAACTGGG GTTTGGGGTG CAAATATCTG CACTGAGTCA  
3651 CTAAATTGCA CCCAGCCTCA TTCCTTTATC TGTAAGTGG GCTAAGAAATG  
3701 CTCCCCTGCC TTCCTCCTCG GTGTAGTACG AGGAAGGATC CCATGACACC  
3751 TGCTCTCCCA GTTTAAAGCT CTATATGTAT GTTGTGAAAT TGACAGGGAT  
3801 CGCTGCACAA ACGCTAATGC AAAGTGGGCT CCTGTGCTTC CTTTTCTCTT  
3851 TCTTCTTCTT TTTTTTTTTT TTAATTTTCT TCTAGAGATG AGGTCTCACT  
3901 ATATTGCCCC GGGTTGGTTT CAAACTCCTA GGGTCAAGCG ATCCTCCAC  
3951 CTTGGCCTCC CAAACTGCTG GTATTACAGG CGTGAGCCAC TCTGTCTGGC  
4001 TCCTATGCTT GTGAATGTCA ACAGCAATCA GCCCTTAGCT GGCAGGGCTG  
4051 GGTGTGGTAGG GCGAGAGCTC ACCCAAGGCT GCTTTTATTA CCCTGCGTGA  
4101 ATCTGCCTGG CCCCTTCCTT CTAAGGAGGT TGCTCTGTGG TTGTCACTCT  
4151 CTCCCTTTAC AGCTGGATCC TGATCTTTCA GTTTCTAACC CTGTCTGAC  
4201 TCATCGTGCT GGAAGTGAGA GCCCGGGGTG AGGTCAGGGA ACTCCCTTGC  
4251 GCGTTTCAAG AAAAGGGAAA AGGAAAGAGA GGTGAGGAGG GGGGCAGATG  
4301 ACCAGAGAGA CACAGGCTGA GAGAGACTGA GACAGACCCA GAGAGCCTCA  
4351 CACATTGAGT GACAGAGACG GAGAAATGGA GATAGGCACC AAAAAATGGT  
4401 TCTCAGTGAC AGAAAGGGAA AAAAGCAACC CCCCAGTCTC TCTTAACATC  
4451 TGGTGAGAAA CCAGCCATGT GCTTTGGTCT GGGCCACAC AGCAAAGGAT  
4501 TATGTAGGGT TTCATGCTGG TGGATGGTCA CCTTATAGCA ACAGGTATCT  
4551 GGGGCTGTCG GGAACACAGA CACGAGGTTG TGGGACCCAG ACCCACAGAG  
4601 ATGGAGCTGT TCTAGGAGCT CTGGTCCCTG TTCTGGTCCC CTGGGATATG  
4651 GCACAGTGAA GGCCACCATC AGGCAGCTGG AGCCAGCAG CAACTGGGAG  
4701 GCAGTAAACA GGGACCGAAA GTGCAAGGTT ACCTCCGAGG CAAACTACTC  
4751 TAAGCTACCC TGTGCTGAGC TCAAGTCCCT TGGAACTATC CCTAAGGCTT  
4801 CCGCTTCCAG AGTGTTTGAG TATTTTCTGT GCACAGCTTC GAATAAATCC  
4851 CACAGCAACA GGTAAACGGC TGCAAGCTGT GACTGTTTTT TAAGAGCTCA  
4901 TCTCACAAAT TCAGGTCCCT TTCATTTAAA CAGAGATGGC AGGAAAGGCG  
4951 TTATTTTGGG ATCTGCATGG AGGAAGTTCA CCAGGCAGCC TCAATTCACC  
5001 AGCTGGAAGT TTGCGTTGTT TGGAAATTTG ATGTGTAACA CGTCTGCAT  
5051 GTGGGCTGAT GTTTTTGTAA ACGGGTAGCA CACACATTCA GCAGGGCACC  
5101 AAAGAGCGGG GGCTTTGCAG TTAGTTCCTT CCTTGGCTCT GCAGCCTTGT  
5151 GTAAGACATG ACACGACTTT GAACCTCTGT TTCCTCTTCT GTGCAAAGCA  
5201 ATGATGACAG TATCTACATC ACAGGACTGG CATGAGGACC AAGTGAGATT  
5251 GGGCAAGGTG CCCGGGCACA CCAGTCTCAC TGTCAGTCTGATGGGCAGA  
5301 GTGGTTGCCCT GGCAGTAGCA TCCTCTATCT TCAGCCACC ACCTCTCTTG  
5351 CTGGCTCACT CCAACTGCTC TTTAGAGATA CACGCTTCCC CTCTTTTCTC  
5401 CTCCCAGTGC CTTTCAGTAT GGCTGCATTT CCCCCTGCAA GTTGGTGTGT  
5451 GCTGGGTGGA GGTGGGGGTG AGGACATGTA TTCTCTGGAG AAGGCCCTGG  
5501 TAACGTCAAA GCACCTCTTT GCTGGTGGCC TGGCCCTGTG ACCTCATTTG  
5551 TACCATTTC TTTTCTAAGA AATACCAGAG ATGGACAGCC ATCTGGTAGA  
5601 GAAGTTGGGC CAGCACCTCT TACCTTGGAT GGACCGGCTT TCCCTGGAGC  
5651 ACTTGAACCC CAGCATCTAT GTGGGCCTAC GCCTCTCCAG TCTGCAGGCT  
5701 GGGACCAAGG AAGACCTCTA CCTGCACAGC CTCAGCTTGT GTTACCAGCA  
5751 GTGCCTCCTA GGGTATTGCC ACACCTCTCT TTTCCATGTC TTGCTCCACA  
5801 TACTAAGAGA TGGGAAACTT GGGTACTAGT TTGGGCCTGT CACCACTTTG  
5851 TGGGCAGACC TTAGGCAAAT TTTCTCCATC TATAGAATGG AGGACCTTTG  
5901 TCCATCTATA GAATGAAGGG GTTGGTTGGA TTAGATCAGA GATGCTAATG  
5951 CAAGGCTCCT TTTGCTACTA CTGTCCATCA TGTGTCTGAG GCAGACATAA  
6001 CTAATCCGTG ACTATACTCT TTGATGATGA GCCCAGGAGC AGCATCTGAC  
6051 TCTATGCTCC CTTAGTGTGC CTGAGGCAGA TATCACTAAT CGATGACTGC  
6101 AGTCTTCTAC ATTGAGCTTA GAAGCAGCAT CTGACTCTGT ATGCTCCCTT  
6151 CCCATGCATG AGGCAGACAT CAGTAATCCA TGACCGCATT CTTTCATACT  
6201 GAGCCCAGAA GCAGCATCTT TTCTTTTCTT TCCTCTCACT CTGTTGCCCA  
6251 GGCTAGAGTG CAGTGGCACA ATCTTGGCTT GCCCAACCT CCAATTCCCG  
6301 GGTTCAGTG ATTCTCGTGC CTCAGCCACC TGAATAGCTG GGATTACAGG  
6351 CGTGTGCCAC CATGCCAGC TGATTTTGTG ATTTTGGTA GAGATAGGCT  
6401 TTCACCATGT TGGCCAGGCT GGTCTTGAAC TCCTGACCTC AGGTGATCCG  
6451 CCTGTCTTGG CTTCCCAAAG TGTGGGATT ATAGGCATGA GCCACTGCAC  
6501 CAATCCAAAA GCAGCATCTT TGTGCTCCCT TTTCAGAGG CATCACAGAG  
6551 AGGCCTGTTT TGGGGTTTGA ATGAGAGGCG AAGAATCAGC CATGGAGTGC  
6601 CTCTTTCTCA GACTCCCTCT TGAGAAGTGG GTGCAGGGGT GGAGAGAAAA  
6651 GAAGACTAGG CATAGTGGCT CATACTGTA ATCCCAACAT TTTGGGAGGC  
6701 TGAGGCAGGA AGATTGCTTG AGCTCAGGAG TTTGAGACCA GCCTAGGCAA  
6751 CATAGTGAGA CCACATCTCT TAAAAAAG AAAAAGAAAA AAAATGAGCC  
6801 AGGTGTAGTG ACTCATGCC GTGGTCCCA CTTCTCCGGA GGCAAGGTG  
6851 GGAGGATCTT TTGAGGCTGA GAAATCGAGG CTACAGTGAG CCATGGTGGC

FIGURE 3B

6901	ACCACCTGCAC	TCCAGCCTGG	GAGACAGAGA	GACCCCTATCT	CAGTAAAAAA
6951	AAAAAATAAA	AATATGGCTG	GGTGTGGTGG	CTCACGCCCTG	TAATCCCAGC
7001	ACTTTGGGAG	GCCAAGGTAG	GTAGATCACA	TGAGGTTAGG	AGTTCGAAAC
7051	CAGTCTGGCC	AACATAGTGA	AACCCTGTCT	CTACTGAAAA	TACAAAAAAT
7101	TAGCCAAGGG	TGGTGGTGGG	CAACTGTAAT	CCCAGCTACT	TGGGAGGCCG
7151	AGGCAGAAGA	ATCGCTTGAA	CTCGGGAGGC	GGAGGTGCA	GTGAGCTGAG
7201	AACATGCCAC	TGCACCTCCAG	CCTGGGCAAC	AAGAGCGAAA	CTCTGTCTCA
7251	AAGAAAAATA	ATAAATAAAA	TAAAAAATA	AAAAAGGAGG	GGGCATATGG
7301	GTGAAGTATG	GACAAAATAG	TGGGGCAGGC	ACAGATGATC	TGGACACAGG
7351	AGCCCTTGGA	GTTTATTCTT	GAATCTAACT	GTTTCATCTT	ATTAAATATT
7401	TGTGGCATAC	ACCTCACAAAC	AACATAGCCA	ACACACCTCC	TTTGGAGCT
7451	TTTATCGAAG	TTTCCCACTG	TTAAGATTTT	TTCCCGCTTT	GTGATGCGGG
7501	TGGGGTGGGT	GCTGTAAAGCA	GGCTTACGGG	GTGGCAGTTT	CTCACAAAGG
7551	CATTAAGTGG	CCTTGTCTTA	GGTCTGCCTT	CAGCGAGGAT	GACGGTGACT
7601	GCCAGGGCAC	GCCTTCCATG	GGCCAGCTGG	CCCTCTACCT	GCTCGCTCTC
7651	AGAGCCAAC	GTGAGTTTGT	CAGGGGCCAC	AAGGGGGACA	GGCTGGTCTC
7701	ACAGCTCAAA	TGGTTCCTGG	AGGATGAGAA	GAGAGCCATT	GGTGAGCAGA
7751	CACCATCCGC	TGGGGGTGGG	GAGCAGCTGG	GAGGGCTCAT	CAGATGATAT
7801	TGTCCCAATGA	GAATCAGAAC	TTTGGGTTTT	CTCCCAGGC	TCTTTTCCCA
7851	CCATCCATTTC	TGCCCATCTC	ACTGCCTACG	TAGAGGCTCG	AACCTGTCCC
7901	CATAGCCATC	CTTGACCCAG	CTTTTCCCGC	GCTGCACACA	TACTATTGAC
7951	AGGTGTGTTT	CGTGGTTTTT	TGTTTTTTGT	TTGTTTGT	GTTTTGAGTT
8001	GGAGGTTTGC	TCTTGCTGCC	CAGGCTGGAG	TACAATGGCG	CAATCTCAGC
8051	TCACCGCAAT	CTCTGCCTCC	TGGGTTCAAG	CAATTCTCTT	GCCTCAGCCT
8101	CCTGAGTAGC	TGGGATTACA	GGCATGCGCC	ACCACACCCA	GCTAATTTTG
8151	TATTTTATAGT	AGACGTGGGG	TTTCTCCATG	TTGGTCAGGC	TGGTCTCGAA
8201	CTCCTGACCT	CAGGTGATCC	GCTTGCTTGA	GCCTCCGAAA	GTGCTGGGAT
8251	TACAGGCATG	AGCCACTGCG	TTAGGCCAC	TGACAAGCCT	TGTATTGGCT
8301	AGCCACCAAG	ATTGACTTGA	TTATCCACCT	TCGGGACAAC	TGGACAGCCT
8351	GCTTATGACT	TACGCCATAG	TCTGTCTCTA	CTAGCTCTCC	TGCCCTGACT
8401	TGACCCAGCA	TACAACAGCC	AGAGCCAGCC	TTTTCAATAT	AAACCTGATC
8451	TGTCTGGCAC	TGCTTAAACC	CTGCAGGGGC	CTCGCACTGC	TCCATGGCCC
8501	AGCCTGTCTA	CCCTTACCTT	CTGCCAGGC	TCTGCTCATC	CATTCTCTGC
8551	CTCCACACA	CCTGCCCTCT	GTGGGCTCCA	GCCATACCAT	CTCTCAACTC
8601	ATAAGCCAGT	TTTTTCATAC	AGGCTCCCTC	CATCTGGACT	GGCTTCCCTG
8651	CGTGCAGTTC	ACTCTGCTC	TACCTTTGGC	TCTGCCTCCA	CCCATCCTCA
8701	GCCGTCTCCA	GCATTACCTC	CTTGGAGAAT	CCTGCCTTGA	CTTCCCAGCC
8751	ACCCAAATAT	CACCTACTTG	TCTGCATTCT	CGTTGCAATT	GCAGTCCGAT
8801	GAGCAATTGC	TGTGGTTGAG	GCCCGAACTG	CGCAAGTGCC	TGTCTGCCAT
8851	GGGTCTCTCG	TTTCTCTTAA	GCACAGTGCC	TGACACACAG	TGAGACCTCA
8901	GCACGTATGG	GCTGAGGCAA	TGAAGGAATG	AAGGATCCCA	TGACCCAAAA
8951	GAGCCTGTTG	GAAAGTGACG	GCCAGGGTCC	CAGGTGCTGG	CGGGGCTGGC
9001	TGCTGGGTGG	GGGCAGAGAG	GCAACCCCTC	TGTTTTTTTC	CCTCTCAGGG
9051	CATGATCACT	AGGGCCACCC	CCACACTAGC	TACTACCAGT	ATGGCCTGGG
9101	CATTCTGGCC	CTGTGTCTCC	ACCAGAAGCG	GGTCCATGAC	AGCGTGGTGG
9151	ACAAACTTCT	GTATGCTGTG	GAACCTTTCC	ACCAGGGCCA	CCATTCTGTG
9201	GGTGAGTAGG	TCAGACCGTG	CCAAGGCCAG	GCTGGCACTC	CCTCAGTCCC
9251	CAGGTCTGCA	CTGATGACGT	CCATACCCCTG	GCCCCACAC	TCACCTTTCC
9301	TTGGGGCTCC	TCCGAATCAA	GTCTTTTAGG	GACGAATTGG	CGAGGGCTCA
9351	TGGGTGATGC	TCCAGCTGTG	AGCCAGCTTT	GGAGCTGGTA	GGTGGATCTC
9401	TTGAGGCCAG	GAGTTCAAGA	CAACGTGGTG	AAACCCCATC	TCTACTAAAA
9451	ATAAAAAAGT	TAGCCGGGCA	TGGTGGCACA	TGCCGTGAGT	CCCAGCTACT
9501	CGGGAGGCTG	AGGCAGGAGA	ATCACTTGAA	CCTGGGAGGC	GGAGGCTGCA
9551	GTGAGTGGAG	ATCGCACCAC	TGCCCTCCAG	CCTGGGCAAC	AGAGTGAGTG
9601	AGACTCTGTC	TCAAAAAATA	AAAAATAAAA	TAAAACTCCC	CTAGTGATTTC
9651	CAATGTGCGAG	CTAAGTTTGG	AAATAGGTGG	TATGGGGTCA	AGTCCTCTTG
9701	GGCCTCCCTC	CTCCAGTCTT	TCTCCCTAAC	CTCTAGCCCT	CAAGTTGCAG
9751	AGTGATCAGC	CAAAACAGTT	TGCCCAGAAA	TGAGCAGTTT	CCTGGGACAC
9801	AGGATTTTCA	GAGTCCAGAC	AAGGAAAGTC	TTGGGCAGAC	CAGGTTGAGT
9851	TGGTGGCCCTT	AGCTGATCTG	ACCATGTTGC	CCTTCTTCTC	CAAGCCCTCC
9901	TGTGGTTGTC	CATAGCTACA	AGGGCCTGAC	CCTCAAGCCC	CTGCCTGTCC
9951	TGGCCCTTTT	GGCTCTCCAG	CTCATTGCAT	GTTCTGTCCC	CCACTTCAAG
10001	ACACAGCAGC	CATGGCAGGC	TTGGCATTCA	CCTGTCTGAA	GCGCTCAAAC
10051	TTCAACCCCTG	CTCGGAGACA	ACGGATCACC	ATGGCCATCA	GAACAGTGGC
10101	AGAGGAGATC	TTGAAGGCCC	AGACCCCGA	GGGCCACTTT	GGGAATGTCT
10151	ACAGCACCCC	ATTGGCATTG	CAGGTGGGAA	AGAGACCCTG	GAGCCATGGC
10201	CACCTGGGG	AACAGTCGGG	TGGAGTGGTC	AGGTGCTGGA	ACACCTAGCC
10251	CCTCCTGCCC	GGCTGACCTC	CTCTCTCTCT	TCCTCACTCT	ATCACCAGTT
10301	CCTCATGACT	TCCCCCATGC	CTGGGGCAGA	ACTGGGAACA	GCATGTCTCA

FIGURE 3C

10351 AGGCGAGGGT TGCTTTGCTG GCCAGTCTGC AGGATGGAGC CTTCCAGAAT  
10401 GCTCTCATGA TTTCCCAGCT GCTGCCCCTT CTGAACCACA AGACCTACAT  
10451 TGATCTGATC TTCCCGAGCT GTCTGGCACC ACGAGGTAGC CCAACTTTT  
10501 GTGGAAGCAC AGCCCTTTAC AATCTGCTGC GCACCCATTG ACGTCCCAGT  
10551 GAGGGGAGGT TGCTTCATCC TGATTTGCTG AGTCAGCACA AGTTTGTGGG  
10601 TGTGCATGGG ACACAGTAGC CAAAATGTGG TCATAGCTTC TAGAAGCTCA  
10651 CAGTGTGGGG AGGAAGACAG TAAATGGAGA TCCCTGGGCA TATCGCTTGT  
10701 GTGATACCCA GTACAGAAAT GTTTGGATGG ATGGATGGAT GGATGGATGG  
10751 ATGGATGGAT GGATGGATGG ATGAGGAGAG ACACATTTTG GTTAACTCTA  
10801 ATACAACATG ATAAGCCCCA GTAGCAGCAT GATCCAGGCT TTCTCTGAGA  
10851 GAGGGTCTGA GGACGTGACT GGGATTTGCC AATTAAGAAT GGAGAAAGAG  
10901 GCCAGGTGCA GTGACTCATG CCTGTAATCC CAACACTTTG GGAGGCCGAG  
10951 GCGGGTGGCT CACCTGAGGT CAGGAGTTCG AGACCAGCCT GGCTAACATG  
11001 GCGAACTCC ATCTATTAAA AATACAAAA AGTAGCTGGG TGTGGTGGCG  
11051 AGTGTCTGGA AGCCAGCTA AGCTACTCAG GAGGCTGAGG CAAGAGAATC  
11101 ACTTGAACCT CAGAGGTGGA GGTTCGAGTG AGCCAAGATC ATGCCACTGC  
11151 ACTCCAGTCT GGGTGACAGA GTAAGACTAT GTCTCAAAAA AAAAAAAAAA  
11201 AAATGGAGAA GAAGGAAGCT GGACATGGTG GCTCGTGCTT ATAATCCTAG  
11251 CACTCTGGGA AGCTGAGGCA GATGGATTGC CTGAGCCGAG GAGTTTGAGA  
11301 CCAGCCTGGG CAACATGGTG AAACCCTGTC TTTACTAAAA TACGAAAGAT  
11351 TAGCCAGGCA TGGTGGTAGA CACCTATAAT CCCAGCTACT AGGGAGGCTG  
11401 AGCCACAAGA ATCACTTGAA CCTGGGAGAC AGAGGTTGCA GTGAGCCGAG  
11451 ATCGCGCCAT TGCATCCAG CCTGGGCGAC AGTGTGAGAC TCTGTCTCCA  
11501 GAAAAACAA GAATGGATAG AGTGGAGCCA AGAAGAGGCA GGAAGAACAA  
11551 AGACACAGAG GTGCACAGAG TTTGGGGGAA TTTTGAGGAA TGGTCTTGCA  
11601 AAAGAGTGGG ATCTGGGAGA ATGAGTGGGA GTGGAAAGCA GATGAATGAA  
11651 GAGAAGTGA GCGCATCAGG GTAACAGAGA TGCGTTGTGA ACAAATGCAT  
11701 GTTCTAGGAA GAGCCCTCTG GAGTGCTAGG TGCCAGAGAG GTGGGAGGAA  
11751 GGATACTGGA AGCAGAGAAA CCAGTGAGGG GCCTGATCTT GGGTGGTGGG  
11801 GAATGAGGGA CAGGGGAGGC CGGGATGGAA GCCAGGTGGT GGGGAATGAG  
11851 GGACAGGGGA GGCCGGGATG GAAGCCAGGT TTCAGCTGAG CAGGTGGCGG  
11901 TGGCATTGAT GGAGATGAGG ACATGGGGAA GGACAAAGTC CAGGTGTCCT  
11951 TGAGGGAAGA CAAGAAGACA AATAATCCAG GCTCTCTGTC CTCACACCAG  
12001 CTGCCCGCCC CTTTCTTCTT GGCACAGTCA TGTGGAACC AGCTGCTGAG  
12051 ACCATTCTCT AGACCAAGA GATCATCAGT GTCACGCTGC AGGTGCTTAG  
12101 TCTCTTCCCG CCGTACAGAC AGTCCATCTC TGTCTGGGCC GGGTCCACCG  
12151 TGAAGATGT CCTGAAGAAG GCCCATGAGT TAGGAGGATT CACGTGAGAC  
12201 TCCCACCTCC CAGTCTCTAC CCCACCCAAC CTCACATGCC TGATAACAGG  
12251 GTCACAGAAA AGACGGGGAA CAGAGGAGAG GGTTCCTCTG GGAGAGACAC  
12301 TGGCCCTGCT TCTGCTTCTA CCTGCTCAGC TCCTTTCTTG CCCACGGTGT  
12351 TATGGAACA GGGAGCCATA GGCCAGCATT GTCAGTGA GAAGCAGGCTT  
12401 TGGAGGCAGA GCGCCCAAGT TGGAAATCCCA ACTCTAACCA GCTAGGTTCC  
12451 AGGTAGGCAC CCACAATTCA CCGAGGAGAA CAGTTGTGCC CCTTCCCTGC  
12501 AGGGCCAGT TCAGAGATCC AGGAGTTAGT ACACATAGAG ATAGTGGCAT  
12551 GTGCTTTTTA TATGTGCAAG GTCCAGCACA TAGCAAGCGC TCAACACAGC  
12601 GTTGCTTTTA TCAGAGTAAG AACTGTTTTT TGTGTTTTTG TTTGTTTTGT  
12651 TTTAAGAGAC AGGGTCTCAA TCTTATCACC CAGGCTGGAG TGTAAATTGT  
12701 CAATCAGTCT TCACTGCAGT CTCGAACCTC GGGGATGAAG CAACCCCTACT  
12751 GTCCTGCCTC AGCCTCCCAA ATAGCTGAGA CTATAGGCAC GTGCCACACA  
12801 ACCCTGGGTA ATTTTTTTTT TTTTTTTTTT GAGATAGGGT CTCTGTCTGT  
12851 TGCCAGGCTT GGTCTCAAAT TCCTGGCCTC AAACCATCCT CACACCTGAG  
12901 GCGCTCAAAA TATTGGGATT ATAGGTGCGA GCCATCATGC TCAGCCAGAA  
12951 TAATAACTGG TTTTTTTTGT TTTTTTTTTG AGACAGAGTC TCACTCTATT  
13001 ACCCAGGCTC TGGAGGCCCA ACTCGTGTGT GTGTATTTGT TTATTTTTAT  
13051 TTATTTATTT ATTTTCGAGC AGAGCCTCTC TCTTTCACCT AGGCTGGAGT  
13101 GCAGTGGCGC AATCTCGGCT CACTGCAACC TCCGTCTCCT GGGTTCAAGT  
13151 GATTGTCTCT CCTCAGCCTC CTGAGTAGCT GGTGCTACAG GCGCGTGCCA  
13201 CCATGCCAGC CTAATTTTTG TATTTTTAGT AGAGACAGGG TTTTACTATG  
13251 TTGGCCAGCT GGTTTCTAAC TCCTGAACCT GGGTGATCTG CCTGCCTCGG  
13301 CCTCCCAAGT TGTCGCGGAT ACAGGCATGG GCCTCCGTGC CCGGCCATGT  
13351 ATTTATTTAG GCAAGGTCTC TCTCTGTTAT CCAGGCTGAA GTGCAGTGGC  
13401 ACATTCATAG CTCACGTCAG CCTCAAATTA TCCAAGTAAC AGGGACTACA  
13451 GGCATGCACC ACCACACCCA TCTACTTTTT TTTGAGATGG AGTCTCCCTC  
13501 TGTGCGCCAG ACTGGGTTGC AGTGGCACAA TTTACGCTCA TGGCAGCATC  
13551 TACCTCCAG GTTCAAGCGA TTCTCCTTCC TCAGTCTCCC GAGTAGCTGG  
13601 GACTATGGGC ATGCACCACC ATACCTGGCT AATGTTTATA TTTTGAGTAG  
13651 AGATGGAATT TTGCCATTTT GGCCAGGCTG GTCTTGAGCT CTTGACCTCA  
13701 AGTGATATGT CTGCCTCAGN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
13751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3D



17251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
19951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
20001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
20051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
20101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
20151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNCCAAATC
20201	AACGAGTTGC	ATAAATCACT	CCTCTATCTT	CCTTGGGGTG	GAAAGTGGAT
20251	GGGAGTTATA	ATTTGAGTTC	TCTTTTGTCT	TAGTCCATTG	AAGCTGCTAT
20301	TACAAATAC	CATAAACTGG	GTGGCTTATA	AACAGCAGAA	ATGAGGCCGG
20351	GTGCGGTGGC	TCATGCCTAT	AATTCAGCA	CTTTGGGAGG	CCAAGGCAGG
20401	TGGATCACCT	GAGATCAGTA	GTTCAAGACT	AGCCTGACCA	ACATGGTGAA
20451	ACCCTGTCTC	TACTAAAAAT	ACAAAAAATT	AGCTGGGGGT	GGTGGCGGGC
20501	ACCTGTAATC	CCAGCTACTC	AGGAGGCTGA	GGCAGGAGAA	TCGCTTGAAC
20551	CCAGGAGGCG	GAGGTTGCCG	TGAGCTGAGA	TCACGCCATT	GCATTTCAGC
20601	CTGGGCACAA	AGAGTGAAAC	TCCATCTCAA	AATGAAATAA	AATAACAGAA
20651	ATGTATTCTT	TAACAGTTCT	GGAGGTTGGG	TGGGCAGTCC	CAGATCAGGA

FIGURE 3F

20701	CACTGACAGA	TTCAGTGTCT	GATGGGGGCC	CACTTTCTGG	TGTTACCTGC
20751	TGGCTGTGTT	CTCACATGGT	GGAAGGAACA	TGGCAACTTT	CTGGGGCCTT
20801	GTTTTTTAAT	TTAAAAAAA	AAAAATATTT	CCTGGCCCTT	GCCTGCTGAA
20851	GGAACCTCTT	TTATAATGGT	ACTTAAAAAT	TTTTTTTTTT	GAGATGGGGG
20901	TCTCACTCTG	TCACCCACGC	TGAGTGCAGT	ATCACAATCT	CAGCTCACTG
20951	CAACCTCTGC	CTCCCTGGCT	TAAGCGATCC	TCCCACCTCA	GCCTCCTGAG
21001	TACGTGTGAC	CATAGGCCCA	TGGCACAAAG	CCCAGCTAAT	TTTTTGTATT
21051	TTTAGTAGAA	ATGTGGTTTC	ACCATGTTGC	ATAGGCTGGT	CTCGAACTTC
21101	TGAACTCAAG	TGATCTGCCT	GCCTTGGCCT	CCCAAAGTGC	TGGGATTCTA
21151	GGTATGAGCC	ACCCTGCTCG	GCCTATAATG	GCACCTTCCT	ATCCCATTTGA
21201	TGAGGCTCTA	CTCTCATGAC	CTAATCATCT	CCCAAAGGCC	CTAAGGCCTC
21251	CTGATACCAT	CACCTTTGGG	GTTAGGTTTT	AACATATACA	TTTTGGGGGG
21301	ACACAGACAT	TTTAGACCAT	AGCACCTCCA	TTGAAAGGAA	ACATTTCTGA
21351	CACCTGGCTA	TCTCAAAGGG	CCCTTTTCAGT	TCCCTGTCAG	GCTGCATTCC
21401	CACATCAACA	CGTAGAGCAG	CGACACTCAC	TCAGAGGTTA	AATAACTTGT
21451	CCAGAGTCAC	AGCAGTAATG	AATGACAGAG	CTGGGGCTTG	AATCCAGGCG
21501	TCCTCCTAGA	GCCTGGATTG	TGTGTAGTGA	GTGAAAGCTG	ACTCCTGGGA
21551	GACTTCTGCG	TGGTCTGGT	TCTCTCTCCA	GACTGCACTG	CGCAAGTTTC
21601	TCTTCTGTAT	GGTCCCTAGG	GTATTACAAA	GACAGTGCCG	CTGCCTGTCA
21651	GGTGTTTTTA	TTACCAGATG	AGGTCATGGC	CTCAGGAACC	CTGTAGGAAG
21701	CTGAGTTTCAG	AGTCTTTGAG	CAGGCTTTAG	GGAGGTTCCA	GCTTCCCACC
21751	ACCAAGCCCC	AGGTGGATTG	TTACAGACTC	TAGCCTCAGG	GTGGGGGGTC
21801	TGGAAGATGA	GGTTGCGGGG	TGCGATATTC	TGCCCAATTC	GCCCCCTCCT
21851	GCTCAATCTG	TTTCTGCAGG	TATTGCTGAC	TACAGACCCA	AGGATGGAGA
21901	AACCATTGAG	CTGAGGCTGG	TTAGCTGGTA	GCCCCTGAGC	TCCCTCATCC
21951	CAGCAGCCTC	GCACACTCCC	TAGGCTTCTA	CCCTCCCTCC	TGATGTCCCT
22001	GGAACAGGAA	GTCGCCGTGAC	CCTGCTGCCA	CCTCCTGTGC	ACTTTGAGCA
22051	ATGCCCCCTG	GGATCACCCC	AGCCACAAGC	CCTTCGAGGG	CCCTATACCA
22101	TGGCCACCTT	TGGAGCAGAG	AGCCAAGCAT	CTTCCCTGGG	AAGTCTTTCT
22151	GGCCAAGTCT	GGCCAGCCTG	GCCCTGCAGG	TCTCCCATGA	AGGCCACCCC
22201	ATGGTCTGAT	GGGCATGAAG	CATCTCAGAC	TCCTTGGCAA	AAAACGGAGT
22251	CCGCAGGCCG	CAGGTGTTGT	GAAGACCACT	CGTTCTGTGG	TTGGGGTCCT
22301	GCAAGAAGGC	CTCCTCAGCC	CGGGGGCTAT	GGCCCTGACC	CCAGCTCTCC
22351	ACTCTGTCTG	TAGAGTGGCA	GCTCCGAGCT	GGTTGTGGCA	CAGTAGCTGG
22401	GGAGACCTCA	GCAGGGCTGC	TCAGTGCCTG	CCTCTGACAA	AATTAAAGCA
22451	TTGATGGCCT	TGGACCTGTC	TACAGTGGCC	TGGTGCCTCA	TACTCCTCAG
22501	GTGCAGGGGC	AGGGACAAGA	GAAGGGGGAA	GTAACCCCAT	CAGGGAGGAG
22551	TGGAGGGTGC	CTGAGCCGCC	ATGTGGGCAT	TGGGGGAGTG	ATGGGAATGC
22601	CAGCAGTGAT	GACGTTGACT	ACTGACTGAG	CACCCACTAC	TATGACTGAG
22651	CACTCACTCG	CTGATACTA	TCTTGAACCTG	CTCTGTGAGG	TTGTTGATAT
22701	TTTCATTTTT	ATCTGTGCTT	TACAAATCAG	GAAACTGGGA	GCCCGGGCGT
22751	GGTGGCTCAC	GCCTGTAATC	CCAGCACTTT	AGGAGGCCAA	GGCAGGTGGA
22801	TCACAAGGTC	AGGAGTTTGA	GATCAGCCTG	GCCAACATGG	TGAAACTCCA
22851	TCTTTACTAA	AAATACAAAA	AATTAGCCAG	GCATGGTGTT	GCATGCCTGC
22901	ATGCCTGTAA	TCCCAGTTAC	TTGGGAAGCT	GAGGCAGGAG	AATTGCTTGA
22951	ACCCTGGAGG	CGGAGGTTGT	AGTGAGCCGA	GATCACGCCA	TTGCACTCCA
23001	GCTTGGGCAA	GAAGAGAAAC	ACTCTCAAAA	AAAAAAAAAA	ATCAGGAAAC
23051	TGGTGCTCAA	AAAGGAAAAA	TGACTCACCA	AGGTCACAGA	CTAGGCAGTG
23101	ATGCTGGGGG	AACCTGGCTC	AGGGGACACA	GACCTGGCCT	GGGGCAGCCT
23151	TGCAGCTCCT	CCACTAAAAT	ACTGAAAATG	AGGGGCTTCG	ATGATGGTTA
23201	TAATCGTATG	GCAGAGCCCC	AACTCAACTG	GAGCCCTGGG	ACCCAGAAGC
23251	TAGGGTCTCA	CTCCCTGCTT	TTCCACAAGG	CACCATTAGG	GCATCACCCC
23301	AGGCCTCGGC	AGCCACGACG	CAGGGATCCT	GCCTCTCATT	GGTTGGGGGC
23351	TTAGGGGCTC	TGGGCTGCCC	TCTTGAAGAG	GGGGTTCAGC	CCAGCGAGGC
23401	ACCCCTATG	CTGCACCCCA	CCAAGGTTAG	GAAGAGGTCC	TGTCCTCAGT
23451	GGGGCCCTCT	GATGAACAGC	CCATCAGGTC	TGCGTCCACA	TGCCTTGGA
23501	GAGATGGTGA	CATACTCAAA	GTCCTTGAAG	CCGCATATTA	AACCACCTAG
23551	AGCACCATCT	TCAAACATTT	AGGGTCTGAG	AAGATAGGGG	AAGTAAGCAA
23601	TTTAAACAT	TTCTTTATAT	TGGGCCAGGT	GCAATGGCTC	ACGTCTGTAA
23651	TCCCAGGCTC	TTGGGAGGAC	GAGGATCACC	TGAGGTGAGG	AGTTCAAGAT
23701	CAGCCTGGCC	AACATGGAGA	AACCCCATCT	CTACTAAAAA	TACAAAAATT
23751	AGCTCAGGCG	TGGTGATGTG	CACCTGTAAT	CCTAGCTATT	CAGGAGGCTG
23801	AGGCACAAGA	ATTGCTTGAG	TCAATATTGC	ACCACTGCAC	TCCAGCCTGG
23851	GCAACAGCGA	GACTCTTGTC	TCAAAAAAAA	AAAAAGATAT	TTGCTGAAAA
23901	GACCCAGCCT	GCCAAACTCA	GGGGCAGCCA	AGGGAGGTAG	TGAAATGGAA
23951	GTTGGAGCTC	AGCGCTCCCA	CACCTCCACT	GCCCTCAGGC	CTTCTCTGCC
24001	TCTTTCCCAT	CAGTCAGCTG	CTTCTGGGCA	TGGTCTGGC	AGAGACTTGG
24051	CCTCCTTCCA	GTTCAAGCTC	CCTCTTAGAT	TGTGTCCCA	GCCACTGAGT
24101	CTTTGGGACA	CTGGGTGAGA	TGTCTAGTCT	GGCACAATTG	GCAGGAATCC

FIGURE 3G

24151 CAAGAAACAG TGTGAGTGAG GGGACAGTCG TGTGAGTGC CCTCCATCTG  
24201 GGACTGGGAG GCAGGTCTAT GTCAGGCCTG CATTTAGATC TCTAATGGCT  
24251 CCAGACAAGC CCCTTCAGCT CACTAAGCCT GTTTCCTAAC ACAGCTGTGG  
24301 GATGGTGCTT TGGTTTACAT AGCACGCGAT ACCATCATAG ATCACATGGG  
24351 GAAACTGAGG CCCCAGGAGT GATCTGCTGG CACATGCAGT GACAAGAGGA  
24401 GAGGCCCATC TCAGCCTTGC AGCAAGGTTG CCAGAAATCG ATTCTCGCCC  
24451 CCATCCCGTA AAGATAGCTG GGATTACAGG TGTGCACCAC CATGCCAGC  
24501 CTAATTTTTG TATTATTAGT AGAGATGGGG TTTCACCATG TTGTCCAGGC  
24551 TGGTCATGAA CTCCTGACCT CAAGTGATCC ACCCGCTTTG GCCTCCCAAA  
24601 GTGCTGGGAT TACAAGCATG AGCCACAGTG CCTGGCCTGA CCCTGCTCTT  
24651 TTGAAAGACC ATTCCCCCAA ATTCTGTGCA CCTGTGTGCC TTTCTTCTCT  
24701 CTGCCTCCTC TCAGCTCTGC CCCGCTCTCC TCCCTTCTCC TCTGGCAAAT  
24751 CCCACTCATC TCTTGAAGCC CTTCTTCCAG GGAAGCCCTT GATCATGCTG  
24801 CTTTCTCCTG TGGGAGGGAT GAAGGACGTG GCCCACGGAG TTTGTTTTGT  
24851 TTTGTTTTTG TGCTCATGTT TGCTCATGTT GCCCAGGCTG GGGTACAATG  
24901 GTACGATCTC AGCTCACTGC AACCTCTACG TCCCGGGTTC AAGCGGTTCT  
24951 CCTGCCTTAG CCTCCCACTG AGCTGGGATT ACTGGCATGA ACCACCACAC  
25001 CTGGCTAATT TTGTGTTTTT AGTAGAGATG GGGTTTCTTC ATGTTGGTCA  
25051 GGCTGGTCTC GAACCCCAA CCTCAGGTGA TCTGCCTGCC TCGGCTCCC  
25101 AAAGTACTGG GATTACAGGG TTGAGCCACT GTGCCTGGCC CAGGCCACG  
25151 GAGTTTTTAAG AGGCTTCCTG TGGCAGTGGC ATCCAGACGG AGTGCAGAAA  
25201 CTCAAAGTTG AAGGCCAGAA GCTCAGGGAA GGGGAGTGTG GAGTTGAGGA  
25251 GTCTCTTGGC TGCCAGGGCC AGAAACCGAA CTCCAAGCCT TCCACAACA  
25301 GCGGGGTAG AGCATGTAGA ATCAGAGAGG AGGCTGAGCC ATGCAGCCCC  
25351 GAGAAGAGGG GAATGCCACT GAGCCACAGA GACCCAGTGC CACTGCCAGG  
25401 TGTCTCTGCC TCCACTTCCC ATGACCCGGC CTGTCTCTGT ATGCAGGCTT  
25451 CACCTCTCTC CGTTGTACAT TGTACACATT CTAGGTGACA CCAGCAGCTT  
25501 CTGATTCTCA TCTCCATAA CATCAGCCCC CCAGAGAGGG GACAACGTCT  
25551 GAGCTGATAA CATAATAGAT GCCCCTTTCC TGGAGGCCAT GGTATGGTC  
25601 AGCGTGAGGA GGATGAAGCC TGAGCAGGCA GGATCGGGGG TCTAGAGGGG  
25651 AAGGAGGTGG AAGTTGAGAT CACAGACCTG TGGTCAGGTG GCCTGGGAAG  
25701 GGTGAGCA GAATGAGCCC AAAGAGCTTG GAAGGGATT TGTGCTGTG  
25751 GGTGAGCACT GCCTCTCCCC TTAGGGACAA CAGCCACCTC TTCTCTCCCC  
25801 ATTTGCTTTT CCCTTCTGTA GATATGAAAC ACAGGCCTCC TTGTGAGGCC  
25851 CCTACTTAAC CTCCTGTATG GGGAAAGCGG CCGGAGAAAG GGAGTTCTGG  
25901 CAGCTTCTCC GAGACCCCAA CACCCCACTG TTGCAAGGTG AGTCATGGCC  
25951 TGACACTCTG GATGTGTCCC CTACCCCAAG CTTACTCAGC CAAGAGGCTT  
26001 CATCAACTCA CCCCAGCTTT CCCTAGCACC CTCTGGGGCC ACACCTTCAC  
26051 AAAATCACTG ATGCTCAAAG TTGGATATAA TATATTGAAC TGAAGCCTTA  
26101 GCATTTTAT GTCAAGTTACT GTGGAATTTC TAGGAAACCA GACAGATTAC  
26151 AAAAAAAAAA AAAAATAAGA AGAAAATTAA CATCACCTAG GATATACTAC  
26201 CTAGGAATAA CGTCTTTTAT TTTGAGATGG AGTTTCGCTC TTGTGCCCCA  
26251 GGCTGGAGTG CAGCGGTATG ATCTCGGCTC GCTGCAACCT CCGCCTCCTG  
26301 GGTTCAATGT ATTCTTCCAC CTCGGCCTTC CTAGAGCCCA AGTGGTCTGC  
26351 CTGCCTCTGC CTCCCAAAGT TCTGGGATTA CAGGCATGAG CCACCGCACC  
26401 CAGCCAAAAT TACTTAACTT TTCTTCTAGA TACTTTTAA AAATATGGCA  
26451 GTAAGTTTTT CATAAAAAAT GGAGCCATGC TATCCAGTGG AAATTTAATG  
26501 TTGCCACAT GTATAACTTA AAAATTTCAT ATATGTGTAT ACATATATAT  
26551 GAAATATATA TATACAGACA CACATATATA TGTATACATA TATATACACA  
26601 TATATATGTA TACATATATA CACATATATA TGTATACATA TATATACACA  
26651 CATATACACA TATATACACA CACATACATA TATACACACA CATATATACA  
26701 CACATATATA CACACATGCA CACATATATA TGTATACATA TATACACACA  
26751 TGTATACGTA TATATACACA CATATATACA CACATATATA TACACACATA  
26801 TACACACATA CACACACATA TATACACACA TATATACACA CATATATACA  
26851 CACATATATA TGTATACATA TATATACACA CATATATACA CACATACACA  
26901 TACATATATA CACATATACA CATATACACA CACATATACA CACATGTATA  
26951 CATATATATA CACACATGTA TACATATGTA TACACACACA TATATGTATA  
27001 CATATATACA CACATACATA TGTGTACATA TATACACACA TACATATGTA  
27051 TACATATATA CACACAT  
(SEQ ID NO:5)

FIGURE 3H



Isoform 1:

**FEATURES:**

Exon: 2031-2094  
 Intron: 2095-5569  
 Exon: 5570-5762  
 Intron: 5763-7571  
 Exon: 7572-7741  
 Intron: 7742-10000  
 Exon: 10001-10173  
 Intron: 10174-10298  
 Exon: 10299-10485  
 Intron: 10486-12027  
 Exon: 12028-12193  
 Intron: 12194-25821  
 Exon: 25822-25939

**Allelic Variants (SNPs):**

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
921	C	T	Beyond ORF (5')			
1781	C	T	Beyond ORF (5')			
1850	G	A	Beyond ORF (5')			
2839	A	G	Intron			
3730	G	A	Intron			
6631	G	A	Intron			
6945	-	A	Intron			
6952	A	T	Intron			
7457	G	A	Intron			
7830	T	A	Intron			
8089	T	C	Intron			
8551	C	T	Intron			
9269	G	C	Intron			
9362	C	T	Intron			
9782	G	T	Intron			
11493	G	A T	Intron			
12260	A	G	Intron			
13086	T	C	Intron			
13183	T	C	Intron			
21240	C	G	Intron			
21695	A	G	Intron			
22058	C	T	Intron			
22233	C	A	Intron			
22245	C	-	Intron			
22375	C	T	Intron			
23042	A	- T	Intron			
23344	T	C	Intron			
23873	A	-	Intron			
24764	G	T	Intron			
24939	T	C	Intron			
24945	G	A	Intron			
25092	C	T	Intron			
25428	T	G	Intron			
25513	C	T	Intron			
25684	C	T	Intron			
26165	A	-	Beyond ORF (3')			

Context:

DNA

Position

921 TTGGAGATATTTTAAGGTCATAGTGTCTTCACAAATTGAGCTGAAAGGGAAGTGTAGGA  
 TGATCTTGCCCTAACCCCTCTCATCTCACACAGGAAGAAGTATTTTAACTCGAGAGGTAA  
 GTGACCTGGCCAAAGTCACACAGCCACCAGTAACTCGTATACATTGATTCTCCTGT  
 GGGGCTGGGCAGATGAGGAATCTTTTGTCTCTTCCCTGTTTGCAGAGATTTTTTTTGAG  
 GTTACTTTCCGAGTCTGGCAAGTACCCCTGCTTCTGGTAGCTTGTGTCTCGATTCAAT  
 [C,T]

**FIGURE 3I**

TCATTCTTTTTATTTTATTTTATTTTGTGAGACAGGGTCTCACTTTGTCAACCAAGCTGGA  
GTGCAGTGGTGAATCTTGGCTCACTGTAGCCTCCACCTCTTGGGTCAAGCGATCCTCC  
TGCCCTCAGCCCCCAAGTAGCTGGGATTACAGACGTCTGCCACACGCCAGGCTAATTTA  
TGGTTTTTTGTATGTGTTTTTTGTGTTTTTTGTAGAGACAGTGTTCCTCATGTGCCCAG  
GCTGGTCTCCAACCTCTGAGCTCAAGTGATCTGCCCGCTCAGCCTTTCAAAGTGCTAGG

1781 ACAGCTGACTCCAGCAATGCTGCTCACGTGACCACTGCAGCTGCAGCTCCCGTTCCACTC  
CTTGTCTCTGGGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGC  
ACCCGGTCTCTGCAATCCACCACTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGG  
ACAGACTTGACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGA  
ACGTGAATGGCCTCAAGATGTCTCACATGGTCCCACTAGCCCTCCTCCTCCCCTTTGTTCC  
[C, T]  
TACCTCCAGGAGGGCTGCTCTGCCCTTCTTCTCTGTCTTTTGGCCTTATGTTCCTCCG  
CACCACAGGCCCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGG  
AGGATTAAATCAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAAGGAG  
AGCCTCAGCAGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGCTCACTGCTCACCACCT  
GCTGCTGCCATGAGGCACCTTGGGGCTTCTCTTCTCTGCGGGTCTGGGGGCCCTC

1850 GGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGCACCGGTCC  
TGCCAATCCCACTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGGACAGACTTG  
ACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGAACGTGAATG  
GCCTCAAGATGTCTCACATGGTCCCACTAGCCCTCCTCCTCCCCTTTGTTCCCTACCTCCA  
GGAGGGTGTCTGCCCCCTTCTTCTCTGTTCTTTGGCCTTATGTTCCCCGCCACACAG  
[G, A]  
CCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGGAGGATTAAT  
CAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAAGGAGCCCTCAGC  
AGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGCTCACTGCTCACCACCTGCTGCTGCC  
ATGAGGCACCTTGGGGCTTCTCTTCTTCTGCGGGTCTGGGGGCCCTCACTGAGATG  
TGTGGTGAGTAACTCGCTCTATCCTGTGCCTTTTCTCTCGGGTCTTAGTGGGTGG

2839 AACATAGGGAGACCCATCTCTACAAAAATAAAAAAATTAATAAATAGCTGGGCATGG  
GGAAGACTTTCTGAAGACCAAGAGGACACATGGGAGCTGAAACTCGAAGGAAGAAAAGGA  
GCTGGCAGGAAGGAGTGGGGGACACACATTCTAGGCAGCAGGAAGTAGCCTTCGGAGG  
TCCTGCCTGTCTCAGCTCTGTGCCCCAAGGGGTCTCTTGGAGCACAGTCTCCTGGGACCT  
GTCTATGAGTCTGAGCTTAGAGGCTCAGGGCTGCTCCTTCAGACAGGAGGCAGGAAGGCAG  
[A, G]  
CTTTGGGAACCTTTGGGCCGCCACCGCGCTTTTCTCCTCCTCTGCACCTAGGATTACGTT  
GAGCAATACACTTTACCCCCATGGTCTCTTGAGACCTTGGGGAAACCTTGAGAGGTGGG  
TGCAGTCATGTCCAGGTGTCAAGTGAAGAAGTCGAGGGTTGGAGGGGCTGAGTGACCCAC  
TCAGGGTGTCTCCACCTTTTCCAGAGCTTTGCTGAACCTAGTTTTTGAAGCTGAAGCCTC  
GTTTGTTTTCTGTTTTGTTTTTGTGTTGAGAGAGGTTCTCCCTCTGTTGCCAGGCTGGAGT

3730 GACACCTCAGGTCTGGGCCAGGAACCCAGCTCTTGGTTTATGTCCGACAGTCCCCAG  
GGGAGTTCTGGGTTCACCCAGCAAGAGCTCTTCTCCTGGCTGATCTGGTCTCAGCCTT  
GGACAGTTAGTCCATTAACTGACCCACAGGAGCCCCAATCCCTTGGGGTCTGGGGAAT  
CTTGAACCTGGGGTTTGGGGTGCAAAATATCTGCACTGAGTCACTTAATTGCACCCAGCCTC  
ATTCTCTTATCTGTAAGTGGGCTAAGAATGCTCCCTGCCTTCTCCTCGGTGTAGTAC  
[G, A]  
AGGAAGGATCCCATGACACCTGCTCTCCAGTTTAAAGCTCTATATGTATGTTGTGAAAT  
TGACAGGGATCGCTGCACAAACGCTAATGCAAGTGGGCTCCTGTGCTTCTTCTCTT  
TCTTCTTCTTTTTTTTTTTTTTAATTTTCTTCTAGAGATGAGGTCTCACTATATTGCCCA  
GGGTTGGTTTCAAACCTCCTAGGGTCAAGCGATCCTCCACCTTGGCCTCCCAAACCTGCTG  
GTATTACAGGCGTGAGCCACTCTGTCTGGCTCCTATGCTTGTGAATGTCAACAGCAATCA

6631 TGAATAGCTGGGATTACAGGCGTGTGCCACCATGCCAGCTGATTTTTGTATTTTGGTA  
GAGATAGGGTTTTACCATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCCG  
CCTGTCTTGGCTTCCCAAAGTGTGGGATTATAGGCATGAGCCACTGCACCAATCCAAAA  
GCAGCATCTTTGTGCTCCCTTTTCAAGAGGCATCACAGAGAGGCTGTTTTGGGGTTTGA  
ATGAGAGGCGAAGAATCAGCCATGGAGTGCCTCTTCTCAGACTCCCTCTTGAGAAGTGG  
[G, A]  
TGCAGGGGTGGAGAGAAAAGAAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATT  
TTGGGAGGCTGAGGCAGGAAGATTGCTTGAGCTCAGGAGTTTGAGACCAGCCTAGGCAAC  
ATAGTGAGACCACATCTTTAAAAAAAAGAAAAAAGAAAAAATGAGCCAGGTGTAGTGA  
CTCATGCTGTGGTCCCCACTTCTCCGGAGGCAAGGTTGGGAGGATCTTTTGAGGCTGAG  
AAATCGAGGCTACAGTGAGCCATGGTGGCACCCTGCACTCCAGCCTGGGAGACAGAGAG

6945 AGAAAAGAAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATTTTGGGAGGCTGAG  
GCAGGAAGATTGCTTGAGCTCAGGAGTTTGAGACCAGCCTAGGCAACATAGTGAGACCAC  
ATCTCTTAAAAAAAAGAAAAAAGAAAAAATGAGCCAGGTGTAGTGAATCATGCTGTGG

FIGURE 3J

5952 AAGACTAGGATAGCTAGTGGCTCATACCTGTAATCCCAACACTTTTGGGAGGCTAGGCCAGGAA  
GATTGCTTGAGCTCAGGAGTTTGAGACCAGCCTAGGCAACATAGTGAGACCACATCTCTT  
AAAAAAAAGAAAAAGAAAAAATGAGCCAGGTGTAGTGACTCATGCCTGTGGTCCCCAC  
TTCGCCGAGGCAAGGTGGGAGGATCTTTTGAGGCTGAGAAATCGAGGCTACAGTGAGC  
CATGGTGGCACCCTGCACCTCCAGCCTGGGAGACAGAGAGACCCTATCTCAGTAAAAAAA  
[A,T]  
AAAATAAAAATATGGCTGGGTGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC  
CAAGGTAGGTAGATCAATAGAGTTAGGAGTTTCGAACCAGCTGTGGCCACATAGTGA  
CCCTGTCTACTGACAAATACAAAAATAGCCAAGGTGGTGGTGGCAACTGTAAATCC  
CAGCTACTTTGGGAGGCCGAGGCAGAAGAATCGCTTGAACTCGGGAGGCGGAGGTTGCAGT  
GAGCTGAGAACATGCCACTGCACCTCCAGCCTGGGCAACAAGAGCGAAACTCTGTCTCAA

7457 AAGAATCGCTTGAACTCGGGAGGCGGAGGTTGCAGTGAGCTGAGAACATGCCACTGCAC  
CCAGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAGAAAATAAATAAAATAAAAA  
AATAAAAAAGGAGGGGGCATATGGGTGAAGTATGGACAAAATAGTGGGGCAGGCACAGAT  
GATCTGGACACAGGAGCCCTTGGAGTTTATTCTTGAATCTAACTGTTTCATCTTTATTTAA  
TATTTGTGGCATACACCTCACAACAACATAGCCAACACACCTCCTTTTGGAGCTTTTATC  
[G,A]  
AAGTTTCCCAGCTGTTAAGATTTTTTCCCGCTTTGTGATGCGGGTGGGTGGGTGCTGTAA  
GCAGGCTTACGGGTGGCAGTTTCTCACAAGGCATTAACCTGGCCTTGTCTAGGTTGTGC  
CTTCAGCGAGGATGACGCTGAGCTGCCAGGCCAAGCCTTCCATGGGGCAGCTGGCCCTCTA  
CCTGCTCGCTCTCAGGCGCAACTGTGAGTTTGTGTCAGGGGCCACAAGGGGGACAGGCTGGT  
CTCACAGCTCAAATGGTTCTTGAGGATGAGAAGAGAGCCATTGGTGAGCAGACACCATC

7830 GGTGGCAGTTTCTCACAAGGCATTAACCTGGCCTTGTCTAGGTCTGCCTTCAGCGAGGA  
TGACGGTGACTGCCAGGGCAAGCCTTCCATGGGCCAGCTGGCCCTCTACCTGTCTCGCTCT  
CAGAGCCAACTGTGAGTTTGTGTCAGGGGCCACAAGGGGGACAGGCTGGTCTCACAGCTCAA  
ATGGTTCTCGGAGGATGAGAAGAGAGCCATTGGTGAGCAGACACCATCCGCTGGGGGTGG  
GGAGCAGCTGGGAGGGCTCATCAGATGATATTCTCCAATGAGAATCAGAACTTTGGGTTT  
[T,A]  
CTCCCCAGGCGCTTTTCCCACCATCCATTCTGCCCATCTCACTGCCTACGTAGAGGCTCG  
AACCTGTCCCATAGCCATCCTTGACCCAGCTTTTCCCGCGCTGCACACATACTATTGAC  
AGGTGTGTTTCGTGGTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG  
CTTCTGCTGCCAGCTGGAGTACAATGGCGCAATCTCAGCTCACCGCAATCTTGCTCTC  
TGGGTTCAAGCAATTCTCTTGCCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGCGCC

8089 ATCAGATGATATTCTCCAATGAGAATCAGAACTTTGGGTTTTCTCCCCAGGCGCTTTTCC  
CACCATCCATTCTGCCATCTCACTGCCTACGTAGAGGCTCGAACCTGTCCCATAGCCA  
TCCTTGACCCAGCTTTTCCCGCGCTGCACACATACTATTGACAGGTGTGTTTCGTGGTTT  
TTTGTTTTTTGTGTTGTTGTTGTTGTTTGTGAGTTGGAGGTTTGCTCTTGCTGCCAGGCTGG  
AGTACAAATGGCGCAATCTCAGCTCACCGCAATCTCTGCCTCCTGGGTTCAAGCAATTCTC  
[T,C]  
TGCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGCGCCACCACCCAGCTAATTTT  
GTATTTTTAGTAGAGCTGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTCCTGACC  
TCAGGTGATCCGCTTGCTTGACCTCCGAAAGTCTGGGATTACAGGCATGAGCCACTGC  
GTTAGGCCCACTGACAAGCCTTGATTGGCTAGCCACCAAGATTGACTTGATTATCCACC  
TTCGGGACAACCTGGACGCTGCTTATGACTTACGCCATAGTCTGTCTCTACTAGCTCTC

8551 TACAGGCATGAGCCACTGCGTTTAGGCCCCACTGACAAGCCTTGATTATGGCTAGCCACCAAG  
ATTGACTTGATTATCCACCTTCGGGACAACCTGGACAGCCTGTTATGACTTACGCCATAG  
TCTGTCTCTACTAGCTCTCTCGCCTGACTTGACCCAGCATACAAAGCCAGCAGCCG  
TTTTCAATATAAACCTGATCTTGCTGGCACTGCTTAAACCTTGACAGGGGCTCGCACTGC  
TCCATGGGCCAGCCTGTCTACCTTACCTTCTGCCAGGCTCTGCTCATCCATTCTCTGC  
[C,T]  
TCCCACACACCTGCCTCTGTGGGCTCCAGCCATACCATCTCTCAACTCATAGGCCAGTT  
TTTTCATACAGGCTCCCTCCATCTGGACTGGCTTCCCTGCGTGCAGTTCACTCCTGCTCT  
ACCTTTGGCTCTGCCTCCACCCATCTCAGCCGCTCTCCAGCATTAACCTCCTGGAGAATT  
CTGCCCTTGACTTCCCAGGCCCAAAATATCACTACTTGGTCTGCATCTCGTTGCAAAATG  
CAGTCGCTAGGCAATTGCTGTGGTGGAGGCCGGAAGCTGCGCAAGTGCCTGTCTGCCATG

## References

9269 AGGCCAGGGTCCCAGGTGCTGGCGGGGCTGGCTGCTGGGTGGGGGAGAGAGGCAACCCC  
TCTGTTTTTTTCCCTCTCAGGGCATGATCACAAGGGCCACCCCACTAGCTACTACCA  
GTATGGGCTGGGCATTCTGGCCCTGTGTCTCCACCAGAAGCGGGTCCATGACAGCGTGGT  
GGACAAACTTCTGTATGCTGTGGAACCTTTCCACCAGGGCCACCATTTCTGTGGGTGAGTA  
GGTCAGACCGTGCCAAGGCCAGGCTGGCACTCCCTCAGTCCCCAGGTCTGCACTGATGAC  
[G, C]  
TCCATACCCTGGCCCCCACACTCACCTTTCTTGGGGCTCCTCCGAATCAAGTCCTTTAG  
GGACGAATTGGCGAGGGCTCATGGGTGATGCTCCAGCTGTGAGCCAGCTTTGGAGCTGGT  
AGGTGGATCTCTTGAGGCCAGGAGTTCAAGACAACGTGGTGAACCCCATCTCTACTAAA  
AATAAAAAAGTTAGCCGGGCATGGTGGCACATGCCTGTAGTCCCAGCTACTCGGGAGGCT  
GAGGCAGGAGAATCACTTGAACCTGGGAGGCGAGGCTGCAGTGAGTGGAGATCGCACCA

9362 GGGCCACCCCACTAGCTACTACCAGTATGGCCTGGGCATTCTGGCCCTGTGTCTCCA  
CCAGAAGCGGGTCCATGACAGCGTGGTGGACAACTTCTGTATGCTGTGGAACCTTTCCA  
CCAGGGCCACCATTTCTGTGGGTGAGTAGGTGAGACCGTGCCAAGGCCAGGCTGGCACTCC  
CTCAGTCCCCAGGTCTGCACTGATGACGTCCATACCCTGGCCCCCACACTCACCTTTCTT  
TGGGGCTCCTCCGAATCAAGTCCTTTAGGGACGAATTGGCGAGGGCTCATGGGTGATGCT  
[C, T]  
CAGCTGTGAGCCAGCTTTGGAGCTGGTAGGTGGATCTCTTGGGCCAGGAGTTCAAGACA  
ACGTGGTGAACCCCATCTCTACTAAAAATAAAAAAGTTAGCCGGGCATGGTGGCACATG  
CCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGG  
AGGCTGCAGTGAGTGGAGATCGCACCCTGCCCTCCAGCTGGGCAACAGAGTGAGTGAG  
ACTCTGTCTCAAAAAATAAAAAATAAACTCCCTAGTGATTCCAATGTGCAGCT

9782 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCG  
GAGGCTGCAGTGAGTGGAGATCGCACCCTGCCCTCCAGCCTGGGCAACAGAGTGAGTGA  
GACTCTGTCTCAAAAAATAAAAAATAAACTCCCTAGTGATTCCAATGTGCAGC  
TAAGTTTGGAAATAGGTGGTATGGGGTCAAGTCCTCTTGGGCCCTCCCTCCAGTCCTT  
CTCCCTAACCTCTAGCCCTCAAGTTGCAGAGTGATCAGCCAACCCAGTTTGGCCAGAAAT  
[G, T]  
AGCAGTTTCTCGGCACACAGGATTTTTCAGAGTCCAGACAAGGAAAGTCTTGGGCAGACCA  
GGTTGAGTTGGTGCCTTAGCTGATCTGACCATGTTGCCCTTCTTCTCCAAGCCCTCCTG  
TGTTGTCCATAGCTACAAGGGCCTGACCCTCAAGCCCTGCCTGTCTGGCCCTTTGG  
CTCTCCAGCTCATTGCATGTTTGTGTCCTCCCTCAAGACACAGCAGCCATGGCAGGCTT  
GGCATTACCTGTCTGAAGCGCTCAAACTTCAACCTGGTGGGAGACAACGGATCACCAT

11493 AAAAAAAAAAATGGAGAAGAAGGAGCTGGACATGGTGGCTCGTGCTTATAATCCTAGCA  
CTCTGGGAAGCTGAGGCAGATGGATTGCCTGAGCCAGGAGTTTGAGACCAGCCTGGGCA  
ACATGGTGAAACCCCTGTCTTTACTAAAAATACGAAAGATTAGCCAGGCATGGTGGTAGACA  
CCTATAATCCCACTACTAGGGAGGCTGAGCCACAAGAATCACTTGAACCTGGGAGACAG  
AGGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCCAGCCTGGGCGACAGTGTGAGACTC  
[G, A, T]  
GTCTCCAGAAAAACAAGAATGGATAGAGTGGAGCCAAGAAGAGGCAGGAAGAACAAGA  
CACAGAGTGCACAGAGTTTGGGGGAATTTTGAAGGAATGGTCTTGCAAAAGAGTGGGATC  
TGGGAGAATGAGTGGGAGTGGAAAGCAGATGAATGAAGAGAAGGTGAGCGCATCAGGGTA  
ACAGAGATGCGTTGTGAACAAATGCATGTTCTAGGAAGAGCCCTCTGGAGTGTAGGTGC  
CAGAGAGGTGGGAGGAAGGATACTGGAAGCAGAGAAACCAGTGAGGGGCTGATCTTGGG

12260 ACAAGAAGACAAATAATCCAGGCTCTCTGTCTCACACCAGCTGCCCCCCCCCTTTCTTCC  
TGGCACAGTCATGTTGGAACAGCTGCTGAGACCATTCTCAGACCCAAGAGATCATCAG  
GTGCACGCTGCAGGTGCTTAGTCTCTTGCCGCCGTACAGACAGTCCATCTCTGTTCTGGC  
CGGGTCCACCGTGAAGATGTCTTGAAGAAGGCCCATGAGTTAGGAGGATTACGTGAGA  
CTCCACCTCCCAGTCTCACCCACCCAACCTCACATGCCTGATAACAGGGTCACAGAA  
[A, G]  
AGACGGGGAACAGAGGAGAGGGTTCCCTCGGGAGAGACACTGGCCCTGCTTCTGCTTCTA  
CCTGCTCAGCTCCTTTCTTGCCACGGTGTATGGAAACAGGGAGCCATAGGCCAGCATT  
GTCACTGAGAGAGCAGGCTTTGGAGGCAGAGCCCCCAGTTGGAATCCCAACTCTAACCA  
GCTAGGTTCCAGGTAGGCACCCACAATTACCGAGGAGAACAGTTGTGCCCCCTTCCCTGC  
AGGGCCAGTGTGAAGAGTCCAGGAGTTAGTACACATAGAGATAGTGGCATGTGCTTTTTA

13086 GGCACGTGCCACACAACCTGGGTAATTTTTTTTTTTTTTTTTTTTGTAGATAGGGTCTCTG  
TCTGTTGCCAGGCTGGTCTCAAATTCTGGCCCTCAAACCATCTCACACCTGAGGCGCT  
CAAAATATTGGGATTATAGGTGCGAGCCATCATGCTCAGCCAGAATAAATACTGGTTTTT  
TTTTTTTTTTTTTTGAGACAGAGTCTCACTCTATTACCCAGGCTCTGGAGGCCCAACTCG  
TGTTTGTGTATTTGTTTATTTTATTTATTTATTTATTTTCGAGACAGAGCCTCTCTTTT  
[T, C]  
ACCTAGGCTGGAGTGAGTGGCGCAATCTCGGCTCACTGCAACCTCCGTCTCCTGGGTTC  
AAGTGATTGTCCTTGCCCTCAGCTCAGTGGTGCTACAGGCGCGTGCCACCATGC  
CCAGCTAATTTTTGTATTTTTAGTAGAGACAGGTTTTTACTATGTTGGCCAGCTGGTTTC

FIGURE 3L

TAACTCCTGAACTCGGGTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGC  
 ATGGGCCTCCGTGCCCCGCCATGTATTTATTTAGGCAAGGTCTCTCTCTGTATCCAGGC

13183 ACCATCCTCACACCTGAGGCGCTCAAAATATTGGGATTATAGGTGCGAGCCATCATGCTC  
 AGCCAGAATAAATACTGGTTTTTTTGTTTTTTTTGGAGACAGAGTCTCACTCTATTAC  
 CCAGGCTCTGGAGGCCCAACTCGTGTGTGTATTGTATTATTTATTTATTTATTTAT  
 TTCGAGACAGAGCCTCTCTCTTTACCTAGGCTGGAGTGCAGTGGCGCAATCTCGGCTCA  
 CTGCAACCTCCGTCTCTGGGTTCAAGTGATTGTCTGCCTCAGCCTCTGAGTAGCTGG  
 [T, C]  
 GCTACAGGCGCGTGCCACCATGCCAGCTAATTTTTGTATTTTAGTAGAGACAGGGTTT  
 TACTATGTGGCCAGCTGGTTTCTAACTCCTGAACTCGGGTGATCTGCCTGCCTCGGCCT  
 CCAAAGTGCTGGGATTACAGGCATGGGCCTCCGTGCCCCGCCATGTATTTATTTAGGCA  
 AGGTCTCTCTCTGTATTATCCAGGCTGAAGTGCAGTGGCACATTATAGCTCACTGCAGCCT  
 CAAATTATCCAAGTAACAGGGACTACAGGCATGCACCACCACCCATCTACTTTTTTTT

21240 TCAGCTCACTGCAACCTCTGCCTCCCTGGCTTAAGCGATCCTCCACCTCAGCCTCCTGA  
 GTACGTGTGACCATAGGCCCATGGCACAAAGCCAGCTAATTTTTTGTATTTTAGTAGA  
 AATGTGGTTTACCATGTGTCATAGGCTGGTCTCGAACTTCTGAACTCAAGTGATCTGCC  
 TGCCTTGGCCTCCCAAAGTGCTGGGATTCTAGGTATGAGCCACCCTGCTCGGCCTATAAT  
 GGCACCTTCTCTATCCCATGTATGAGGCTCTACTCTCATGACCTAATCATCTCCCAAAGGC  
 [C, G]  
 CTAAGGCTCCTGATACCATCACCTTTGGGGTTAGGTTTAAACATATACATTTTGGGGGG  
 ACACAGACATTTTAGCCATAGCACCTCCATTGAAAGGAAACATTTCTGACACCTGGCTA  
 TCTCAAAGGGCCTTTTCAAGTTCCCTGCAGGCTGCATTCCACATACCAACAAGAGCAG  
 CGACACTCACTCAGAGTTAAATAACTTGTCCAGAGTCACAGCAGTAATGAATGACAGAG  
 CTGGGGCTTGAATCCAGGCGTCTCCTAGAGCCTGGATTCTGTGTAGTGAGTGAAAGCTG

21695 CATTTCCACATACCAACAAGAGCAGCGACACTCACTCAGAGGTTAAATAACTTGTCCAG  
 AGTCACAGCAGTAATGAATGACAGAGCTGGGGCTTGAATCCAGGCGTCTCTAGAGCCT  
 GGATTCTGTGTAGTGAGTGAAAGCTGACTCCTGGGAGACTTCTGCGTGGTCTGGTTCTC  
 TCTCCAGACTGCACCTGCGCAAGTTTCTCTTCTGTATGGTCCCTAGGGTATTACAAAGACA  
 GTGGCCTTGCCTGTGAGGTGTTTTTATTACCAGATGAGGTATGGCCTCAGGAACCTGT  
 [A, G]  
 GGAAGCTGAGTTCAAGTCTTTGAGCAGGCTTTAGGGAGGTTCCAGCTTCCCACCACCAA  
 GCCCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGTCTGGAAGATGAGGTTG  
 CGGGGTGCGATATTCTGCCCAATTGCCCCCTCCTTGCTCAATCTGTTTCTGAGGTATTG  
 CTGACTACAGACCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCC  
 TGAGCTCCTCATCCCAGCAGCTCGCACACTCCCTAGGCTTCTACCTCCTCCTGTATG

22058 CCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGTCTGGAAGATGAGGTTGCG  
 GGGTGCGATATTCTGCCCAATTGCCCCCTCCTTGCTCAATCTGTTTCTGCAAGTATTGCT  
 GACTACAGACCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCCTG  
 AGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCTCCTCCTGTATGTC  
 CTTGGAACAGGAACCTCGCCTGACCCTGTGACCTCCTGTGCACTTTGAGCAATGCCCC  
 [C, T]  
 TGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGGAGCAG  
 AGAGCCAAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCA  
 GGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGC  
 AAAAAACGGAGTCCGAGGCCGAGGTGTTGTGAAGACCCTCGTTCTGTGGTTGGGGTC  
 CTGCAAGAAGGCTCCTCAGCCCGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCT

22233 CCCTGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCTCCTCCTG  
 ATGTCCCTGGAACAGGAACCTCGCCTGACCTGCTGCCACCTCCTGTGCACTTTGAGCAAT  
 GCCCCCTGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTG  
 GAGCAGAGAGCCAAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGC  
 CCTGCAGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTC  
 [C, A]  
 TTGGCAAAAAACGGAGTCCGAGGCCGAGGTGTTGTGAAGACCCTCGTTCTGTGGTTG  
 GGGTCTTGCAAGAAGGCTCCTCAGCCCGGGGCTATGGCCCTGACCCAGCTCTCCACT  
 CTGCTGTTAGAGTGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGAGACCTCAGCA  
 GGGCTGCTCAGTGCTGCTCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTAC  
 AGTGGCCTGGTGCTCATACTCCTCAGGTGCAGGGGACGGGACAAGAGAAGGGGAAGTA

22245 TCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCTCCTCCTGATGTCCCTGGAA  
 CAGGAACCTCGCCTGACCTGCTGCCACCTCCTGTGCACTTTGAGCAATGCCCCCTGGGAT  
 CACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGAGCAGAGAGCC  
 AAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTC  
 CATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGCAAAAAA  
 [C, -]

FIGURE 3M

GGAGTCCGCAGGCCGCAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAG  
AAGGCCTCCTCAGCCCGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTAGAG  
TGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGT  
GCCTGCCTCTGACAAAATTAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTG  
CCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGG

22375 ACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGGAGCAGAGAGCCAAGCATCTTC  
CCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTCCCATGAAGGC  
CACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGCAAAAACGGAGTCCGC  
AGGCCGCAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAGAAGGCCTCC  
TCAGCCCGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTAGTAGTGGCAGCTC  
[C, T]  
GAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGTGCCTGCCTCT  
GACAAAATTAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTGCCTCATACTC  
CTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGGAGGAGTGGAG  
GGTGCCTGAGCCGCCATGTGGGCATTGGGGGAGTGTGGGAATGCCAGCAGTGTACGT  
TGACTACTGACTGAGCACCCACTACTATGACTGAGCACTCACTCGCTAGATACTATCTTG

23042 GCCGGGCTGGTGGCTCACGCCTGTAATCCAGCACTTTAGGAGGCCAAGGCAGGTGGAT  
CACAAGGTGAGGAGTTTGTAGATCAGCCTGGCCAACATGGTGAAGTCCATCTTTACTAAA  
AATACAAAAAATTAGCCAGGCATGGTGTGCATGCCTGCATGCCTGTAATCCAGTTACT  
TGGGAAGCTGAGGCAGGAGAATTGCTTGAACCTTGGAGGCGGAGGTGTAGTGAGCCGAG  
ATCAGCCATTGCACTCCAGCTTGGGCCAAGAAGAGAAACACTCTCAAAAAAAAAAAAAA  
[A, -, T]  
CAGGAACTGGTGTCTCAAAAAGGAAAAGTGACTCACCAAGGTACAGACTAGGCAGTGAT  
GCTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCCTGGGGCAGCCTTGCAGTCTCTCC  
ACTAAAACTACTGAAAAATGAGGGCTTCGATGATGTTATAATCGTATGGCAGAGCCCCAA  
CTCAACTGGAGCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCA  
CCATTAGGGCATCACCCAGGCCTCGGCAGCCACGACGAGGGATCCTGCCTCTCATTGG

25344 AGGAACTGGTGTCTCAAAAAGGAAAAGTGACTACCAAGGTACAGACTAGGCAGTGATG  
CTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCCTGGGGCAGCCTTGCAGTCTCTCCA  
CTAAAATACTGAAAATGAGGGGCTTCGATGATGTTATAATCGTATGGCAGAGCCCCAAC  
TCAACTGGAGCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCAC  
CATTAGGGCATCACCCAGGCCTCGGCAGCCACGACGAGGGATCCTGCCTCTCATTGGT  
[T, C]  
GGGGGCTTAGGGGCTCTGGGCTGCCCTCTTGAAGAGGGGTTACGCCAGCGAGGCACCC  
CCTATGCTGCACCCACCAAGGTTAGGAAGAGGTCTCTGCTCAGTGGGGCCCTCTGATG  
AACAGCCCATCAGGTCTCGTCCACATGCCTTGAAGAGATGGTGACATACTCAAAGTCC  
TTGAAGCCGCATATTAAACCACTAGAGCACCATCTTCAAACATTTAGGGTCTGAGAAGA  
TAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGCAATGGCTCACGT

23873 GGTCTGAGAAGATAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGC  
AATGGCTCACGTCTGTAATCCCAGCGCTTTGGGAGGACGAGGATCACCTGAGGTGAGGAG  
TTCAAGATCAGCCTGGCCAACATGGAGAAACCCATCTCTACTAAAAATACAAAATTAG  
CTCAGGCCTGGTGTGTCACCTGTAATCCTAGCTATTGAGGAGGCTGAGGCACAAGAAT  
TGCTTGAGTCAATATTGCACCACTGCACTCCAGCCTGGGCAACAGCGAGACTCTTGCTCTC  
[A, -]  
AAAAAAAAAAAGATATTGCTGAAAAGACCCAGCCTGCCAAACTCAGGGGCAGCCAAGG  
GAGGTAGTGAAATGGAAGTTGGAGCTCAGCGCTCCACACCTCCACTGCCCTCAGGCCTT  
CTCTGCCCTCTTTCCCATCAGTCAGCTGCTTCTGGGCATGGTCTGGCAGAGACTTGGCCT  
CCTTCCAGTTCAAGCTCCCTCTTAGATTGTGTCCACGCCACTGAGTCTTTGGGACACTG  
GGTCAGATGTCTAGTCTGGCACAAATTGGCAGGAATCCCAAGAAACAGTGTGAGTGAGGGG

24764 ATAGCTGGGATTACAGGTGTGCACCACCATGCCAGCCTAATTTTTGTATTATTAGTAGA  
GATGGGGTTTACCATGTTGTCCAGGCTGGTCACTGAACCTTGACCTCAAGTGATCCACC  
CGCTTTGGCCTCCCAAAGTGTGGGATTACAAGCATGAGCCACAGTGCCTGGCCTGACCC  
TGCTCTTTTGAAAGACCATTCGCCCAAATCTGTGCACCTGTGTGCCCTTCTTCTCTCTG  
CCTCCTCTCAGCTCTGCCCGCTCTCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCT  
[G, T]  
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGTGCTTTCTCCTGTGGGAGGGATGAAG  
GACGTGGGCCACGGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATGTTGCC  
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC  
GGTCTCCTGCTTACCTTCCAGTACCTGAGTGGGATTACTGGCATGAACCACCACTGG  
CTAATTTTGTGTTTTTAGTAGAGATGGGTTTCTTCATGTTGGTCAGGCTGGTCTCGAAC

24939 GACCTGCTCTTTTGAAAGACCATTCGCCCAAATCTGTGCACCTGTGTGCCTTTCTTCT  
CTCTGCCTCTCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCA  
TCTCTGAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGTGCTTTCTCCTGTGGGAGGG

FIGURE 3N

ATGAAGGACGTGGCCACGGAGTTTGTGTTTGTGTTTGTGATGGAGTTTGTCTCATG  
 TTGCCACGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGT  
 [T, C]  
 CAAGCGGTTTCTCTGCTTACGCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACA  
 CCTGGCTAATTTTGTGTTTGTAGTAGAGATGGGGTTCTTCATGTTGGTCAGGCTGGTCT  
 CGAACTCCCAACCTCAGGTGATCTGCCTGCCTCGGCCCTCCCAAAGTACTGGGATTACAGG  
 GTTGAGCCACTGTGCCTGGCCAGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGG  
 CATCCAGACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTG

24945 GCTCTTTTGAAGACCATTCCCCCAAATTCGTGCACCTGTGTGCCTTCTCTCTCTGC  
 CTCCTCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCTT  
 GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTCTCTCTGTGGGAGGGATGAAG  
 GACGTGCCCCACGGAGTTTGTGTTTGTGTTTGTGATGGAGTTTGTCTCATGTTGCC  
 AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC  
 [G, A]  
 GTTCTCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACACCTGGC  
 TAATTTTGTGTTTGTAGTAGAGATGGGGTTCTTCATGTTGGTCAGGCTGGTCTCGAACT  
 CCCAACCTCAGGTGATCTGCCTGCCTCGGCCCTCCCAAAGTACTGGGATTACAGGGTTGAG  
 CCACGTGCCTGGCCAGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGCATCCA  
 GACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTT

25092 ATCATGTCTGCTTCTCTCTGTGGGAGGGATGAAGGACGTGGCCACGGAGTTTGTGTT  
 TTGTTTGTGAGATGGAGTTTGTCTCATGTTGCCAGGCTGGGGTACAATGGTACGATCTCA  
 GCTCACTGCAACCTCTACGTCCCGGGTTCAAGCGGTTCTCTGCCTTAGCCTCCCCAGTA  
 GCTGGGATTACTGGCATGAACCACCACCTGGCTAATTTTGTGTTTGTAGTAGAGATGG  
 GGTTTCTTCATGTTGGTCAGGCTGGTCTCGAACTCCCAACCTCAGGTGATCTGCCTGCCT  
 [C, T]  
 GGCTCCCAAAGTACTGGGATTACAGGGTTGAGCCACTGTGCCTGGCCAGGCCACGGA  
 GTTTTAAAGAGGCTTCCGTGGCAGTGGCATCCAGACGGAGTGCAGAACTCAAAGTTGAA  
 GGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG  
 AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG  
 GCTGAGCATGACGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA

25428 AGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGA  
 GTCTCTTGGCTGCCAGGGCCAGAAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAG  
 AGCATGTAGAATCAGAGAGGAGGCTGAGCCATGACGCCCCGAGAAGAGGGGAATGCCACT  
 GAGCCACAGAGACCCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCC  
 [T, G]  
 GCCTGTCTCTGTATGCAGGCTTCACCCCTCTCTCGTTGTACATTGTACACATTCTAGGTGA  
 CACCAGCAGTTCTGATTCTCATCTCCCATAAACATCAGCCCCCAGAGAGGGGACAACTG  
 CTGAGCTGATAACATAATAGATGCCCCTTTCTTGGAGGCCATGGTCATGGTCAGCGTGGA  
 GAGGATGAAGCCTGAGCAGGCAGGATCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTT

25513 GGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG  
 AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG  
 GCTGAGCCATGACGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA  
 CTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATGCAGGCTTCA  
 CCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCGAGCTTCTGATTCTCATC  
 [C, T]  
 CCCATAACATCAGCCCCCAGAGAGGGGACAACTGCTGAGCTGATAACATAATAGATGCC  
 CCTTCTCTGGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGAGCAGGCAGGA  
 TCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGGTCAGGTGGCC  
 TGGGAAGGGTTTGACGAGTGTGGCCCAAAGAGCTTGAAGGGATTGCTGTCTGTGGGT  
 GAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTCTCTCCCCATTGCTTTCCC

25684 CCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATG  
 CAGGCTTCACCCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTG  
 ATTCTCATCTCCCATAAACATCAGCCCCCAGAGAGGGGACAACTGCTGAGCTGATAACAT  
 AATAGATGCCCCTTTCTTGGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGA  
 GCAGGCAGGATCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGG  
 [C, T]  
 CAGGTGGCCTGGGAAGGGTTTGACGAGTGTGGCCCAAAGAGCTTGAAGGGATTGCTGTCT  
 GCTGTGGGTGAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTTCTCTCCCCATTT  
 GCCTTTCCCTTCTGTAGATATGAAACACAGGCCCTCCTTGTGAGGCCCTTACTTAACCTCC  
 GTGATGGGGAAAGCGCCGGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC  
 CCACTGTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA

26165 GTGATGGGGAAGCGGCCGGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC  
 CCACCTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA

FIGURE 30

CTCAGCCAAGAGGCTTCATCAACTCACCCAGCTTTCCTAGCACCCCTCCTGGGCCACAC  
CTTCACAAAATCACTGATGCTCAAAGTTGGATATAATATATTGAAGTGAAGCCTTAGCAT  
TTTTATGCAAGTTACTGTGGAAATTCTAGGAAACCAGACAGATTACAAAAAAAAAAAAA  
[A, -]  
CTAGAAGAAAATTAACATCACCTAGGATATACTACCTAGGAATAACGTCTTTTATTTGA  
GATGGAGTTTCGCTCTTGTGTTGCCCAGGCTGGAGTGCAGCGGTATGATCTCGGCTCGCTGC  
AACCTCCGCCTCCTGGGTTCAATGTGATTCTTCCACCTCGGCCTTCCTAGAGCCCAAGTGG  
TCTGCCTGCCTCTGCCTCCCAAAGTTCTGGGATTACAGGCATGAGCCACCGCAGCCAGCC  
AAAATTACTTAACTTTTCTTCTAGATACTTTTAAAAATATGGCAGTAAGTTTTCATAA

FIGURE 3P



Isoform 2:

**FEATURES:**

Exon: 2132-2195  
Intron: 2196-5670  
Exon: 5671-5863  
Intron: 5864-7672  
Exon: 7673-7761  
Intron: 7762-9149  
Exon: 9150-9302  
Intron: 9303-10101  
Exon: 10102-10274  
Intron: 10275-10399  
Exon: 10400-10586  
Intron: 10587-12128  
Exon: 12129-12294  
Intron: 12295-25922  
Exon: 25923-26040

**Allelic Variants (SNPs):**

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
1022	C	T	Beyond ORF (5')			
1882	C	T	Beyond ORF (5')			
1951	G	A	Beyond ORF (5')			
2940	A	G	Intron			
3831	G	A	Intron			
6732	G	A	Intron			
7558	G	A	Intron			
7931	T	A	Intron			
8190	T	C	Intron			
8652	C	T	Intron			
9370	G	C	Intron			
9463	C	T	Intron			
9883	G	T	Intron			
11594	G	A T	Intron			
12361	A	G	Intron			
13187	T	C	Intron			
13284	T	C	Intron			
21341	C	G	Intron			
21796	A	G	Intron			
22159	C	T	Intron			
22334	C	A	Intron			
22346	C	-	Intron			
22476	C	T	Intron			
23143	A	- T	Intron			
23445	T	C	Intron			
23974	A	-	Intron			
24865	G	T	Intron			
25040	T	C	Intron			
25046	G	A	Intron			
25193	C	T	Intron			
25529	T	G	Intron			
25614	C	T	Intron			
25785	C	T	Intron			
26266	A	-	Beyond ORF (3')			

Context:

DNA

Position

1022 TTGGAGATATTTTAAGGTCATAGTGTCTTCACAAATTGAGCTGAAAGGGAACGTTAGGA  
TGATCTTGCCCTAACCCCTCATCTCACACAGGAAGAACTATTTTAAACTCGAGAGGTTAA  
GTGACCTGGCCAAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCCTGT  
GGGGCTGGGCAGATGAGGAATCTTTTGTCTCTTCCCTGTTTGCAGAGATTTTGTGAG  
GTTACTTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTTGTGTCTCGATTCAAT

**FIGURE 3Q**

[C, T]  
 TCATTCTTTTATTTATTTATTTATTTTGTAGACAGGGTCTCACTTTGTCAACCAAGCTGGA  
 GTGCAGTGGTGTAAATCTTGGCTCACTGTAGCCTCCACCTCTTGGGTCAAGCGATCCTCC  
 TGCCTCAGCCCCCAAGTAGCTGGGATTACAGACGTCTGCCACCACGCCAGGCTAATTTA  
 TGGTTTTTGTATGTGTTTTTTGTGTTTTGTAGAGACAGTGTTCCTCCATGTGCCCAG  
 GCTGGTCTCCAACCTCTGAGCTCAAGTGATCTGCCCGCTCAGCCTTTCAAAGTGCTAGG

1882 ACAGCTGACTCCAGCAATGCTGCTCACGTGACCACTGCAGCTGCAGCTCCCGTTCCACTC  
 CTTGTCTCTGGGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGC  
 ACCCGGTCCTGCCAATCCACCCTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGG  
 ACAGACTTGACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGA  
 ACGTGAATGGCCTCAAGATGTCTCACATGGTCCCACTAGCCCTCCTCTCCCTTTGTTC  
 [C, T]  
 TACCTCCAGGAGGGCTGCTCTGCCCTTCCTTCTCTGTTCCTTTGGCCTTATGTTCCCGC  
 CACCACAGGCCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGG  
 AGGATTAAATCAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAG  
 AGCCTCAGCAGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGCTCACTGCTCACCACCT  
 GCTGCTGCCATGAGGCACCTTGGGGCCTTCCTTCTCTCTGCGGGTCTGCGGGCCCTC

1951 GGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGCACCGGTCC  
 TGCCAATCCACCCTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGGACAGACTTG  
 ACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGAACGTGAATG  
 GCCTCAAGATGTCTCACATGGTCCCACTAGCCCTCCTCTCCCTTTGTTCCTTACCTCCA  
 GGAGGGCTGCTCTGCCCTTCCTTCTCTGTTCCTTTGGCCTTATGTTCCCCGCCACCACAG  
 [G, A]  
 CCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGGAGGATTAAT  
 CAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAGAGCCTCAGC  
 AGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGCTCACTGCTCACCACCTGTGCTGCC  
 ATGAGGCACCTTGGGGCCTTCTCTTCTCTTCTGCGGGTCTGCGGGCCCTCACTGAGATG  
 TGTGGTGAGTAACTCGCCTCTATCCTGTGCCTCTTTCCTCTGGGTCTTAGTGGGGTGG

2940 AACATAGGGAGACCCCATCTCTACAAAAATAAAAAAATTAATAATAGCTGGGCATGG  
 GGAAGACTTTCTGAAGACCAAGAGGACACATGGGAGCTGAAACTCGAAGGAAGAAAAGGA  
 GCTGGCAGGAAAGGAGTGGGGGACACACATTCTAGGCAGCAGGAAGTGAAGCTTCGGAGG  
 TCCTGCGCTGCTCCAGCTCTGTGCCCAAGGGGTCTCTTGGAGCACAGTCTCTTGGGACCT  
 GTCTATAGTCTGAGCTTAGAGGCTCAGGGCTGCTCCTTCAGACAGGAGGCAGAAGGCAG  
 [A, G]  
 CTTTGGGAACCTTGGGCGCCCCACGCGCCTTTTCTCTCTCTGCACCTAGGATTACGTT  
 GAGCAATACACTTTACCCCCATGGTCTCTTGAGACCTGGGGAAACCTGAGAGGTGGG  
 TGCAGTCATGTCCAGGTGTCAAGTGAAGAAAGTCGAGGGTGGAGGGGGCTGAGTGACCCAC  
 TCAGGGTGTCTCCACCTTTTCCAGAGCTTTGCTGAACTTAGTTTTTAGAACTTGAAGCCTC  
 GTTGTGTTCGTTTGTTTTTTGTGTGAGAGAGGTCTCCCTCTGTTGCCAGGCTGGAGT

3831 GACACCTCAGGTCTGGGCCCAGGAACCCAGCTCTTGGTTTATGTCCGGACAGTCCCCAG  
 GGGAGTTCTGGGTTCAACCAGCAAGAGCTCTTCTCTCTGGCTGATCTGGTCTCAGCCTT  
 GGACAGTTAGTCCATTAACTGACCCACAGGAGCCCCAATCCCTTGGGGTCTGGGGAAT  
 CTTGAACTGGGGTTTGGGGTGCAAATATCTGCACTGAGTCACTTAATTGCACCCAGCCTC  
 ATTCTTTTATCTGTAAAGTGGGCTAAGAATGCTCCCTGCCTCTCTCGGTGTAGTAC  
 [G, A]  
 AGGAAGGATCCCATGACACCTGCTCTCCAGTTTAAAGCTCTATATGTATGTTGTGAAAT  
 TGACAGGATCGCTGCACAAACGCTAATGCAAAGTGGGCTCCTGTGCTTCCTTTTCTCTT  
 TCTTCTTCTTTTTTTTTTTTTTAAATTTTCTTCTAGAGATGAGGTCTCACTATATTGCCCA  
 GGGTTGGTTTCAAACCTCTAGGGTCAAGCGATCCTCCACCTTGGCCTCCCAAACCTGCTG  
 GTATTACAGGCGTGAGCCACTCTGTCTGGCTCCTATGCTTGTGAATGTCAACAGCAATCA

6732 TGAATAGCTGGGATTACAGGCGTGTGCCACCATGCCAGCTGATTTTTGTATTTTGGTA  
 GAGATAGGGTTTCAACATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCCG  
 CCTGTCTTGGCTTCCCAAAGTGTGGGATTATAGGCATGAGCCACTGCACCAATCCAAAA  
 GCAGCATCTTTGTGCTCCCTTTTCAAGAGGCATCACAGAGAGGCTGTTTTGGGGTTTGA  
 ATGAGAGGCGAAGAAATCAGCCATGGAGTGCCTTTTCTCAGACTCCCTCTTGAGAAGTGG  
 [G, A]  
 TGCAGGGGTGGAGAGAAAAGAAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATT  
 TTGGGAGGCTGAGGCAGGAAGATTGCTTGAGCTCAGGAGTTTGAGACCAGCCTAGGCAAC  
 ATAGTGAGACCACATCTCTTAAAAAAAAGAAAAAGAAAAAATGAGCCAGGTGTAGTGA  
 CTCATGCCTGTGGTCCCCACTTCTCCGAGGCAAGGTGGGAGGATCTTTTGAGGCTGAG  
 AAATCGAGGCTACAGTGAGCCATGGTGGCACCCTGCACTCCAGCCTGGGAGACAGAGAG

7558 AAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTGAGAACATGCCACTGCACT  
 CCAGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAGAAAATAAATAAATAAATAAAAA

FIGURE 3R

AATAAAAAAGGAGGGGCATATGGGTGAAGTATGGACAAAATAGTGGGGCAGGCACAGAT  
GATCTGGACACAGGAGCCCTTGGAGTTTATTCTTGAATCTAACTGTTTCATCTTTATTTAAA  
TATTTTGGCATACCTTCACAACAACATAGCCAACACACCTCCTTTTGGAGCTTTTATC  
[G, A]  
AAGTTTCCCACTGTTAAGATT'TTTTCCCGCTTTGTGATGCGGGTGGGTGGGTGCTGTAA  
GCAGGCTTACGGGGTGGCAGTTTCTCACAAAGGCATTAAGTGGCCTTGTCCTAGGCTGTC  
CTTCAGCGAGGATGACGGTGACTGCCAGGGCAAGCCTTCCATGGGCCAGCTGGCCCTCTA  
CCTGCTCGCTCTCAGAGCCAACGTGTAGTTTGTTCAGGGGCCACAAGGGGGACAGGCTGGT  
CTCACAGCTCAAATGGTTCCTGGAGGATGAGAAGAGAGCCATTGGTGAGCAGACACCATC

7931 GGTGGCAGTTTCTCACAAAGGCATTAAGTGGCCTTGTCCTAGGCTGCTTCAGCGAGGA  
TGACGGTGACTGCCAGGGCAAGCCTTCCATGGGCCAGCTGGCCCTCTACCTGCTCGCTCT  
CAGAGCCAACGTGTAGTTTGTTCAGGGGCCACAAGGGGGACAGGCTGGTCTCACAGCTCAA  
ATGGTTCTCTGGAGGATGAGAAGAGAGCCATTGGTGAGCAGACACCATCCGCTGGGGGTGG  
GGAGCAGCTGGGAGGGCTCATCAGATGATATTCTCCAATGAGAATCAGAACTTTGGGTTT  
[T, A]  
CTCCCCAGGCGCTTTTCCACCATTCTGCCCATCTCACTGCCTACGTAGAGGCTCG  
AACCTGTCCCCATAGCCATCTTGACCCAGCTTTTCCCGCGCTGCACACATACTATTGAC  
AGGTGTGTTTCGTGGTTT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT  
TCTTGCTGCCCAGGCTGGAGTACAATGGCGCAATCTCAGCTCACCGCAATCTCTGCCTCC  
TGGGTTCAAGCAATTCTCTTGCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGCGCC

8190 ATCAGATGATATTCTCCAATGAGAATCAGAACTTTGGGTTTCTCCCCAGGCGTCTTTCC  
CACCATCCATTCTGCCCATCTCACTGCCTACGTAGAGGCTCGAACCTGTCCCCATAGCCA  
TCCTTGACCCAGCTTTTCCCGCGCTGCACACATACTATTGACAGGTGT'TTTCGTGGTTT  
TTTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT'TTGT  
AGTACAATGGCGCAATCTCAGCTCACCGCAATCTCTGCCTCCTGGGTTCAAGCAATTCTC  
[T, C]  
TGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGCGCCACCACACCCAGCTAATTTT  
GTATTTT'TAGTAGACGTGGGGTTTCTCCATGTTGGTCAAGGCTGGTCTCGAACTCCTGACC  
TCAGGTGATCCGCTTGCTTAGCCTCCGAAAGTGTGGGATTACAGGCATGAGCCACTGC  
GTTAGGGCCACTGACAAGCCTTGTATTGGCTAGCCACCAAGATGACTTGATTATCCACC  
TTCCGGGACAACTGGACAGCCTGCTTATGACTTACGCCATAGTCTGTCTCTACTAGCTCTC

8652 TACAGGCATGAGCCACTGCGTTAGGCCCCACTGACAAGCCTTGTATTGGCTAGCCACCAAG  
ATTGACTTGATTATCCACCTTCGGGACAACCTGGACAGCCTGCTTATGACTTACGCCATAG  
TCTGTCTCTACTAGCTCTCCTGCCCTGACTTGACCCAGCATAACAAGCCAGAGCCAGCC  
TTTTCATATAAACCTGATCTTGCTGGCACTGCTTAAACCCTGCAGGGGCTCGCACTGC  
TCCATGGCCAGCCTGTCTACCTTACCTTCTGCCCAGGCTCTGCTCATCCATTCTCTGC  
[C, T]  
TCCACACACCTGCCCTCTGTGGGCTCCAGCCATACCATCTCTCAACTCATAAGCCAGTT  
TTTTCATACAGGCTCCCTCCATCTGGACTGGCTTCCCTGCGTGCAGTTCACTCCTGCTCT  
ACCTTTGGCTCTGCCTCCACCATCTCAGCCGCTCTCCAGCATTACCTCCTTGGAGAATC  
CTGCCTTGACTTCCAGCCACCCAAATATCACTACTTGGTCTGCTATTCTCGTTGCAATTG  
CAGTCGCATGAGCAATTGCTGTGGTTGAGGCCCCGAAGTGCCTGTCTGCCATG

9370 AGGCCAGGGTCCCAGGTGCTGGCGGGGCTGGCTGCTGGGTGGGGGCAGAGAGGCAACCCC  
TCTGTTT'TTTTCCCTCTCAGGGCATGATCACAAGGGCCACCCCACTAGCTACTACCA  
GTATGGCCTGGGCATTCTGGCCCTGTGTCTCCACCAGAAGCGGGTCCATGACAGCGTGGT  
GGACAAACTTCTGTATGCTGTGAACCTTTCACCAGGGCCACCATCTGTGGGTGAGTA  
GGTCAGACCGTGCCAAGGCCAGGCTGGCACTCCCTCAGTCCCCAGGCTGCACTGATGAC  
[G, C]  
TCCATACCCTGGCCCCCACACTCACCTTTCCTTGGGGCTCCTCCGAATCAAGTCTTTAG  
GGACGAATTGGCGAGGGCTCATGGGTGATGCTCCAGCTGTGAGCCAGCTTTGGAGCTGGT  
AGGTGGATCTCTTGAGGCCAGGAGTTCAAGACAACGTGGTGAACCCCATCTCTACTAAA  
AATAAAAAAGTTAGCCGGGCATGGTGGCACATGCCTGTAGTCCAGCTACTCGGGAGGCT  
GAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGCTGCAGTGAGTGAGATCGCACCA

9463 GGGCCACCCCCACACTAGCTACTACCAGTATGGCCTGGGCATTCTGGCCCTGTGTCTCCA  
CCAGAAGCGGGTCCATGACAGCGTGGTGGACAACTTCTGTATGCTGTGGAACCTTTCCA  
CCAGGGCCACCATCTGTGGGTGAGTAGGTGAGCCGTGCCAAGGCCAGGCTGGCACTCC  
CTCAGTCCCCAGGTCTGCACTGATGACGTCCATACCCTGGCCCCCACACTCACCTTTCCT  
TGGGGCTCCTCCGAATCAAGTCTTTAGGGACGAATTGGCGAGGGCTCATGGGTGATGCT  
[C, T]  
CAGCTGTGAGCCAGCTTTGGAGCTGGTAGGTGGATCTCTTGAGGCCAGGAGTTCAAGACA  
ACGTGGTGAAACCCCATCTCTACTAAAAATAAAAAAGTTAGCCGGGCATGGTGGCACATG  
CCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGG  
AGGCTGCAGTGAGTGAGATCGCACCACTGCCCTCCAGCCTGGGCAACAGAGTGAGTGAG  
ACTCTGTCTCAAAAAATAAAAAATAAAAACTCCCTAGTGATTCCAATGTGCAGCT

FIGURE 3S

9883 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCG  
GAGGCTCAGTGAGTGGAGATCGCACCCTGCCCTCCAGCCTGGGCAACAGAGTGAGTGA  
GACTCTGTCTCAAAAAATAAAAAATAAAATAAACTCCCCTAGTGATTCCAATGTGCAGC  
TAAGTTTGGAAATAGGTGGTATGGGGTCAAGTCTCTTTGGGCCTCCCTCCTCCAGTCCCT  
CTCCCATAACCTCTAGCCCTCAAGTTGCAGAGTGATCAGCCAAACCAGTTTGCCCAAGAAAT  
[G, T]  
AGCAGTTTCTGGGACACAGGATTTTCAGAGTCCAGACAAGGAAAGTCTTGGGCAGACCA  
GGTTGAGTTGGTGCCCTTAGCTGATCTGACCATGTTGCCCTTCTTCTCCAAGCCCTCCTG  
TGGTTGTCCATAGCTACAAGGGCTGACCCCTCAAGCCCTGCTGTCTGGCCCTTTGG  
CTCTCCAGCTCATTGATGTTCTGTCCCCCACTTCAAGACACAGCAGCCATGGCAGGCTT  
GGCATTACCTGTCTGAAGCGCTCAAACTTCAACCCTGGTCGGAGACAACGGATCACCAT

11594 AAAAAAAAAAATGGAGAAGAAGGAAGCTGGACATGGTGGCTCGTGCTTATAATCCTAGCA  
CTCTGGGAAGCTGAGGCAGATGGATTGCCTGAGCCCAAGAGTTTGAGACCAGCCTGGGCA  
ACATGGTGAAACCCTGTCTTTACTAAAAATACGAAAGATTAGCCAGGCATGGTGGTAGACA  
CCTATAATCCCAGCTACTAGGGAGGCTGAGCCACAAGAAATCACTTGAACCTGGGAGACAG  
AGGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCCAGCCTGGGCGACAGTGTGAGACTC  
[G, A, T]  
GTCTCCAGAAAAACAAGAATGGATAGAGTGGAGCCAAGAAGAGGCAGGAAGAACAAGA  
CACAGAGGTGCACAGAGTTTGGGGGAATTTTGAGGAATGGTCTTGCAAAAGAGTGGGATC  
TGGGAGAATGAGTGGGAGTGGAAAGCAGATGAATGAAGAGAAGGTGAGCGCATCAGGGTA  
ACAGAGATGCGTTGTGAACAAATGCATGTTCTAGGAAGAGCCCTCTGGAGTGCTAGGTGC  
CAGAGAGGTGGGAGGAAGGATACTGGAAGCAGAGAAACCAGTGAGGGGCTGATCTTGGG

12361 ACAAGAAGACAAATAATCCAGGCTCTCTGTCTCACACCAGCTGCCCGCCCTTTCTTCC  
TGCCACAGTCATGTTGGAACCAAGCTGCTGAGACCATTCTCAGACCCCAAGAGATCATCAG  
TGTCACGCTGCAGGTGCTTAGTCTCTTGCCGCGGTACAGACAGTCCATCTCTGTCTTGGC  
CGGGTCCACCGTGGAAAGATGTCTGAAGAAGGCCATGAGTTAGGAGGATTACGTGAGA  
CTCCCACCTCCCAGTCTCACCCCAACCTCACATGCCTGATAACAGGGTCACAGAA  
[A, G]  
AGACGGGGAACAGAGGAGAGGGTTCCCTCGGGAGAGACTGGCCCTGCTTCTGCTTCTA  
CCTGCTCAGCTCCTTTCTTGCCACGGTGTATGGAACAGGGAGCCATAGGCCAGCATT  
GTCAGTGAGAGCAGGCTTTGGAGGCAGAGCCCCCAGTTGGAATCCCAACTTAACCA  
GCTAGGTTCCAGTAGGCACCAATTCACCGAGGAGAAGTGTGCCCCCTCCCTGC  
AGGGCCAGTGCAGAGTCCAGGAGTTAGTACATAGAGATAGTGGCATGTGCTTTTTTA

13187 GGCACGTGCCACACAACCTGGGTAATTTTTTTTTTTTTTTTTTTTGGAGATAGGGTCTCTG  
TCTGTTGCCAGGCTGGTCTCAAATTCCTGGCCTCAACCATCCTCACACCTGAGGCGCT  
CAAAATATGGGATATAGGTGCGAGCCATCATGCTCAGCCAGAATAATAACTGGTTTTT  
TTTGTTTTTTTTTTGGAGACAGAGTCTCACTCTATTACCCAGGCTCTGGAGGCCCAACTCG  
TGTTTGTGTATTTGTTTATTTTATTTATTTATTTATTTTCGAGACAGAGCCTCTCTCTT  
[T, C]  
ACCTAGGCTGGAGTGAGTGGCGCAATCTCGGCTCACTGCAACCTCCGTCTCTGGGTTT  
AAGTGATTGTCTGCCTCAGCCTCCTGAGTAGCTGGTGCTACAGGCGCGTGCCACCATGC  
CCAGCTAATTTTTGTATTTTAGTAGAGACAGGGTTTTACTATGTTGGCCAGCTGGTTTT  
TAACTCCTGAACCTCGGGTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGC  
ATGGCCCTCCGTGCCCGGCCATGTATTTATTTAGGCAAGGTCTCTCTCTGTTATCCAGGC

13284 ACCATCCTCACACCTGAGGCGCTCAAAATATTGGGATTATAGGTGCGAGCCATCATGCTC  
AGCCAGAATAAATACTGGTTTTTTTTTTTGGTTTTTTTTTGGAGACAGAGTCTCACTCTATTAC  
CCAGGCTCTGGAGGCCCAACTCGTGTGTTGTGATTTGTTTATTTTATTTATTTATTTAT  
TTTCAGACAGAGCCTCTCTCTTTACCTAGGCTGGAGTGCAAGTGGCGCAATCTCGGCTCA  
CTGCAACCTCCGTCTCTGGGTTCAAGTGATTGTCTGCCTCAGCCTCCTGAGTAGCTGG  
[T, C]  
GCTACAGGCGCGTGCCACCATGCCAGCTAATTTTTGTATTTTAGTAGAGACAGGGTTT  
TACTATGTTGGCCAGCTGGTTTCTAACTCCTGAACTCGGGTGATCTGCCTGCCTCGGCCT  
CCCAAAGTGCTGGGATTACAGGCATGGGCCTCCGTGCCCGGCCATGTATTTATTTAGGCA  
AGGTCTCTCTCTGTTATCCAGGCTGAAGTGAGTGACATTACATAGCTCACTGCAGCCT  
CAAATTATCCAAGTAACAGGGACTACAGGCATGCACCACCAACCCATCTACTTTTTTTT

21341 TCAGCTCACTGCAACCTCTGCCTCCCTGGCTTAAGCGATCCTCCACCTCAGCCTCCTGA  
GTACGTGTGACCATAGGCCCATGGCACAAAGCCAGCTAATTTTTTGTATTTTGTAGTAGA  
AATGTGGTTTACCATGTTGCAATAGGCTGGTCTCGAACTTCTGAACTCAAGTGATCTGCC  
TGCCCTGGCCCTCCCAAAGTGCTGGGATTCTAGGTATGAGCCACCCTGCTCGGCCTATAAT  
GGCACTTTCCTATCCCATGATGAGGCTCTACTCTCATGACCTAATCATCTCCCAAAGGC  
[C, G]  
CTAAGGCCTCCTGATACCATCACCTTTGGGGTTAGGTTTTAACATATACATTTTGGGGGG  
ACACAGACATTTTAGACCATAGCACCTCCATTGAAAGGAAACATTTCTGACACCTGGCTA

FIGURE 3T

TCTCAAAGGGCCCTTTTCAGTTCCCCTGCAGGCTGCATTCCACATCACCAACAAGAGCAG  
CGACACTCACTCAGAGTTAAATAACTTGTCCAGAGTCACAGCAGTAATGAATGACAGAG  
CTGGGGCTTGAATCCAGGCGTCTCTAGAGCCTGGATTCTGTGTAGTGAGTGAAAGCTG

21796 CATTCCCACATCACCAACAAGAGCAGCGACACTCACTCAGAGGTTAAATAACTTGTCCAG  
AGTCACAGCAGTAATGAATGACAGAGCTGGGGCTTGAATCCAGGCGTCTCTAGAGCCT  
GGATTCTGTGTAGTGAGTGAAAGCTGACTCCTGGGAGACTTCTGCGTGGTCTGTGGTTCTC  
TCTCCAGACTGCAGTGCAGCAAGTTTCTCTTCCCTGATGGTCCCTAGGGTATTACAAAGACA  
GTGGCCCTGCCTGTCAAGTGTTTTATTACCAGATGAGGTATGGCCCTCAGGAACCCCTGT  
[A, G]  
GGAAGCTGAGTTCAGAGTCTTTGAGCAGGCTTTAGGGAGGTTCAGCTTCCCACCACCAA  
GCCCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGGTCTGGAAGATGAGGTG  
CGGGGTGCGATATTCTGCCCAATTCGCCCTCTTGTCAATCTGTTCTGCAAGTATTG  
CTGACTACAGACCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCC  
TGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCTCCCTCCTGATG

22159 CCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGGTCTGGAAGATGAGGTTCG  
GGGTGCGATATTCTGCCCAATTCGCCCTCTTGTCTCAATCTGTTTCTGCAGGTATTGCT  
GACTACAGACCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCCTG  
AGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCTCCCTCCTGATGTC  
CCTGGAACAGGAACTCGCCTGACCCTGTGCCACCTCCTGTGCACTTTGAGCAATGCCCT  
[C, T]  
TGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGGAGCAG  
AGAGCCAAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCA  
GGTCTCCCATGAAGGCCACCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGC  
AAAAAACGGAGTCCGCGAGGCCGAGGTGTTGTGAAGACCACCTCGTTCTGTGGTTGGGGTC  
CTGCAAGAAGGCCCTCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCT

22334 CCCTGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCTCCCTCCTG  
ATGTCCTTGGAAACAGGAACTCGCCTGACCCTGTGCCACCTCCTGTGCACTTTGAGCAAT  
GCCCCCTGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTG  
GAGCAGAGAGCCAAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGC  
CCTGCAGGTCTCCCATGAAGGCCACCCATGGTCTGATGGGCATGAAGCATCTCAGACTC  
[C, A]  
TTGGCAAAAAACGGAGTCCGCGAGGCCGAGGTGTTGTGAAGACCACCTCGTTCTGTGGTTG  
GGGTCTCTGCAAGAAGGCCCTCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACT  
CTGCTGTTAGAGTGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCA  
GGGCTGCTCAGTGCCTGCTCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTAC  
AGTGGCCTGGTGCCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTA

22346 TCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCTCCCTCCTGATGTCCCTGGAA  
CAGGAACCTCGCCTGACCTGTGCCACCTCCTGTGCACTTTGAGCAATGCCCTTGGGAT  
CACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGGAGCAGAGAGCC  
AAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTC  
CCATGAAGGCCACCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGCAAAAA  
[C, -]  
GGAGTCCGCGAGGCCGAGGTGTTGTGAAGACCACCTCGTTCTGTGGTTGGGGTCTGCAAG  
AAGGCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTTAGAG  
TGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGT  
GCCTGCCTCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTG  
CCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGG

22476 ACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGGAGCAGAGAGCCAAGCATCTTC  
CCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTCCCATGAAGGC  
CACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGCAAAAAACGGAGTCCGC  
AGGCCGAGGTGTTGTGAAGACCACCTCGTTCTGTGGTTGGGGTCTGCAAGAAGGCCCTCC  
TCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTTAGAGTGGCAGCTC  
[C, T]  
GAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGTGCCTGCCTCT  
GACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTGCCTCATACTC  
CTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGGAGGAGTGGAG  
GGTGCCTGAGCCGCCATGTGGGCATTGGGGAGTGATGGGAATGCCAGCAGTGATGACGT  
TGACTACTGACTGAGCACCCACTACTATGACTGAGCACTCACTCGCTAGATACTATCTTG

23143 GCCGGCGTGGTGGCTCACGCCGTGAATCCCAGCACTTTAGGAGGCCAAGGCAGGTGGAT  
CACAAGGTGAGGAGTTTGTAGATCAGCCTGGCCAACATGGTGAAACTCCATCTTTACTAAA  
AATACAAAAAATTAGCCAGGCATGGTGTGCATGCCTGCATGCCTGTAATCCCAGTTACT  
TGGGAAGCTGAGGCAGGAGAATTGCTTGAACCCCTGGAGGCGGAGGTTGTAGTGAGCCGAG  
ATCACGCCATTGCACTCCAGCTTGGGCAAGAAGAGAACTCTCAAAAAAAAAAAAAA

FIGURE 3U

[A, -, T]  
CAGGAAACTGGTGCTCAAAAAGGAAAAGTGA CTACCAAGGTCACAGACTAGGCAGTGAT  
GCTGGGGGAACCTGGGCTCAGGGGACACAGACCTGGCCTGGGGCAGCCTTGCAGCTCCTCC  
ACTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAA  
CTCAACTGGAGCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCA  
CCATTAGGGCATCACCCAGGCCTCGGCAGCCACGACGAGGGATCCTGCCTCTCATTGG

23445 AGGAAACTGGTGCTCAAAAAGGAAAAGTGA CTACCAAGGTCACAGACTAGGCAGTGATG  
CTGGGGGAACCTGGGCTCAGGGGACACAGACCTGGCCTGGGGCAGCCTTGCAGCTCCTCCA  
CTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAAC  
TCAACTGGAGCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCA  
CATTAGGGCATCACCCAGGCCTCGGCAGCCACGACGAGGGATCCTGCCTCTCATTGGT  
[T, C]  
GGGGGCTTAGGGGCTCTGGGCTGCCCTCTTGAAGAGGGGGTTAGCCAGCGAGGCACCC  
CCTATGCTGCACCCCAAGGTTAGGAAGAGGTCTGTCTCAGTGGGGCCCTCTGATG  
AACAGCCCATCAGGTCTGCGTCCACATGCCTTGAAGAGATGGTGACATACTCAAAGTCC  
TTGAAGCCGCATATTAAACCACTAGAGCACCATCTTCAAACATTTAGGGTCTGAGAAGA  
TAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGCAATGGCTCACGT

23974 GGTCTGAGAAGATAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGC  
AATGGCTCACGCTCTGTAATCCCAGCGCTTTGGGAGGACGAGGATCACCTGAGGTGAGGAG  
TTCAAGATCAGCCTGGCCAACATGGAGAAACCCCATCTCTACTAAAAATACAAAAATTAG  
CTCAGGCGTGGTGATGTCACCTGTAATCCTAGCTATTCAAGAGGCTGAGGCACAGAAT  
TGCTTGAGTCAATATTGCACCACTGCACTCCAGCCTGGGCAACAGCGAGACTCTTGCTCTC  
[A, -]  
AAAAAAAAAAAAAGATATTTGCTGAAAAGACCCAGCCTGCCAAACTCAGGGGCAGCCAAGG  
GAGGTAGTGAAAAGTTGGAAGTTCAGGCTCAGCGCTCCCAACCTCCACTGCCCTCAGGCCTT  
CTCTGCCTCTTTCCCATCAGTCAGCTGCTTCTGGGCATGGTCTGGCAGAGACTTGGCCT  
CCTTCCAGTTCAAGCTCCCTCTTAGATTTGTGTCCACGCCACTGAGTCTTTGGGACACTG  
GGTCAGATGTCTAGTCTGGCACAATTGGCAGGAATCCAAGAAACAGTGTGAGTGAGGGG

24865 ATAGCTGGGATTACAGGTGTGCACCACCATGCCAGCCTAATTTTGTATTATTAGTAGA  
GATGGGGTTTACCATGTTGTCCAGGCTGGTCACTGAATCCTGACCTCAAGTGATCCACC  
CGCTTTGGCCTCCCAAAGTGCTGGGATTACAAGCATGAGCCACAGTGCTGGCCTGACCC  
TGCTCTTTTGAAGACCAATCCCCAAATTCGTGACCTGTGTGCCCTTCTTCTCTCTG  
CCTCCTCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTTGGCAATCCCACTCATCTCT  
[G, T]  
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG  
GACGTGGCCCCACGGAGTTTGTTTTGTTTTGTGAGATGGAGTTTGTCTCATGTTGCC  
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC  
GGTTCTCCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACACCTGG  
CTAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCTCGAAC

25040 GACCTGCTCTTTTGAAGACCAATCCCCCAAATTCGTGCACCTGTGTGCCTTTCTTCT  
CTCTGCCTCCTCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCA  
TCTCTTGAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGG  
ATGAAGGACGTGGCCCACGGAGTTTGTTTTGTTTTGTGAGATGGAGTTTGTCTCATG  
TTGCCAGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGT  
[T, C]  
CAAGCGGTTCTCCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACA  
CCTGGCTAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCT  
CGAACTCCCAACCTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAGTACTGGGATTACAGG  
GTTGAGCCACTGTGCCTGGCCCAGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGG  
CATCCAGACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTG

25046 GCTCTTTTGAAGACCAATCCCCCAAATTCGTGCACCTGTGTGCCTTTCTTCTCTG  
CTCCTCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCTT  
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG  
GACGTGGCCCACGGAGTTTGTTTTGTGTTTTGTGAGATGGAGTTTGTCTCATGTTGCC  
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC  
[G, A]  
GTTCTCCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACACCTGGC  
TAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCTCGAACT  
CCCAACCTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAGTACTGGGATTACAGGTTGAG  
CCACTGTGCTGGCCCAGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGGCATCCA  
GACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTT

25193 ATCATGCTGCTTTCTCCTGTGGGAGGGATGAAGGACGTGGCCCACGGAGTTTGTGTTGTT  
TTGTTTTGAGATGGAGTTTGTCTCATGTTGCCAGGCTGGGGTACAATGGTACGATCTCA

FIGURE 3V

GCTCACTGCAACCTCTACGTCCCGGGTTCAAGCGGTTCTCCTGCCTTAGCCTCCCCAGTA  
GCTGGGATTACTGGCATGAACCACCACACCTGGCTAATTTTGTGTTTTAGTAGAGATGG  
GGTTTTCTTCATGTTGGTCAGGCTGGTCTCGAACTCCCAACCTCAGGTGATCTGCCTGCCT  
[C, T]  
GGCCTCCCAAAGTACTGGGATTACAGGGTTGAGCCACTGTGCCTGGCCCAGGCCACGGA  
GTTTTAAGAGGCTTCTGTGGCAGTGGCATCCAGACGGAGTGCAGAACTCAAAGTTGAA  
GGCCAGAAGCTCAGGGAAGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG  
AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG  
GCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA

25529 AGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGAGTGTGAGTTGAGGA  
GTCTCTTGGCTGCCAGGGCCAGAAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAG  
AGCATGTAGAATCAGAGAGGAGGCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACT  
GAGCCACAGAGACCCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCC  
[T, G]  
GCCTGTCTCTGTATGCAGGCTTCAACCCTCTCTCGTTGTACATTGTACACATTCTAGGTGA  
CACCAGCAGCTTCTGATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGGACAACCTG  
CTGAGCTGATAACATAATAGATGCCCCCTTCTCGGAGGCCATGGTCATGGTCAGCGTGGA  
GAGGATGAAGCCTGAGCAGGCAGGATCGGGGTCTAGAGGGAAGGAGGTGGAAGTT

25614 GGCCAGAAGCTCAGGGAAGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG  
AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG  
GCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA  
CTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATGCAGGCTTCA  
CCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTGATTCTCATC  
[C, T]  
CCCAATAACATCAGCCCCCAGAGAGGGGACAACCTGCTGAGCTGATAACATAATAGATGCC  
CCTTTCTCGGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGAGCAGGCAGGA  
TCGGGGGTCTAGAGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGGTCAGGTGGCC  
TGGGAAGGGTTTACGAGTGTTCGGCCCAAAGAGCTTGGGAAGGGATTTTGTCTGTGTGGT  
GAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTTCTCTCCCATTTGCCTTTCCC

25785 CCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATG  
CAGGCTTCAACCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTG  
ATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGGACAACCTGCTGAGCTGATAACAT  
AATAGATGCCCCCTTCTCGGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGA  
GCAGGCAGGATCGGGGTCTAGAGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGG  
[C, T]  
CAGGTGGCCTGGGAAGGGTTTACGAGTGTTCGGCCCAAAGAGCTTGGGAAGGGATTTTGTCT  
GCTGTGGGTGAGCACTGCCTCTCCCCCTTAGGGACAACAGCCACCTCTTCTCTCCCATTT  
GCCTTTCCCTTCTGTAGATATGAAACACAGGCCTCCTTGTGAGGCCCTACTTAACCTCC  
GTGATGGGGAAAGCGGCCGGAGAAAGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC  
CCACTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA

26266 GTGATGGGGAAGCGGCCGGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC  
CCACTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA  
CTCAGCCAAGAGGCTTCACTCAACTCACCCAGCTTTCCTAGCACCTCCTGGGCCACAC  
CTTCACAAAATCACTGATGCTCAAAGTTGGATATAATATATTGAACTGAAGCCTTAGCAT  
TTTTATGCAAGTTACTGTGGAATTTCTAGGAAACCAGACAGATTACAAAAAAAAAAAAA  
[A, -]  
CTAGAAGAAAATTAACATCACCTAGGATATACTACCTAGGAATAACGTCTTTTATTTTGA  
GATGGAGTTTCGCTCTTGTGCCCAGGCTGGAGTGCAGCGGTATGATCTCGGCTCGCTGC  
AACCTCCGCCCTCCTGGGTTTCATGTGATTCTTCCACCTCGGCCCTTCTAGAGCCCAAGTGG  
TCTGCCTGCCTCTGCCTCCCAAAGTTCTGGGATTACAGGCATGAGCCACCGCACCCAGCC  
AAAATTACTTAACCTTTCTTCTAGATACTTTTAAAAATATGGCAGTAAGTTTTCATAA

FIGURE 3W